SAMPLED-DATA AND REDUCED ORDER CONTROLLER IMPLEMENTATION

Anton Madievski

MSc, Moscow State University

February 1995

A thesis submitted for the degree of Doctor of Philosophy

of the Australian National University

Department of Systems Engineering

Research School of Information Sciences and Engineering

The Australian National University

Statement of Originality

The contents of this thesis are the results of original research, and have not been submitted for a higher degree at any other university or institution.

Much of the work in this thesis has been published or has been submitted for publication in refereed journals:

- Anton G. Madievski, Brian D.O. Anderson and Michel Gevers, "Optimum FWL Design of Sampled–Data Controllers", *Automatica*, to appear in vol.31, issue 4, 1995.

- Anton G. Madievski and Brian D.O. Anderson, "Sampled–Data Controller Reduction Procedure", *IEEE Trans. Automat. Contr.*, submitted.

- Anton G. Madievski, Brian D.O. Anderson and Yutaka Yamamoto, "Frequency Response of Sampled–Data Systems", under preparation.

- Sang Woo Kim, Brian D.O. Anderson and Anton G. Madievski, "Error Bound for Transfer Function Order Reduction Using Frequency Weighted Balanced Truncation", *Systems & Control Letters*, to appear in vol.24, issue 3, 1995.

- Sang Woo Kim, Brian D.O. Anderson and Anton G. Madievski, "Multiplicative Approximation of Transfer Functions with Frequency Weighting", *Systems & Control Letters*, to appear in vol.24, 1995.

- Venkatappa Sreeram, Brian D.O. Anderson and Anton G. Madievski, "New Results on Frequency Weighted Balanced Reduction Technique", *IEEE Trans. Automat. Contr.*, submitted.

A number of papers have been published in conference proceedings or have been submitted to conferences. Some of the material covered in these papers overlaps with that covered in the publications listed above:

- Anton G. Madievski, "Optimal FWL Design of Sampled-Data Controllers", *Proceedings of the 12th IFAC 1993,* Vol. 5, p.145–148.

- Anton G. Madievski and Brian D.O. Anderson, "A Lifting Technique for Sampled-Data Controller Reduction for Closed-Loop Transfer Function Consideration", *Proceedings of the 32nd IEEE Conf. Decision and Control 1993*, Vol.3, p.2929–2930.

- Anton G. Madievski and Brian D.O. Anderson, "Frequency Response of Sampled-Data Systems", *Proceedings of the 1st Asian Control Conference,* Tokyo, July 27–30, 1994, Vol.1, pp.209–212.

- Venkatappa Sreeram, Brian D.O. Anderson and Anton G. Madievski, "New Results on Frequency Weighted Balanced Reduction Technique", *Proceedings of the 1995 ACC,* to appear.

The research described in this thesis is the result of a collaborative effort with my PhD supervisor Prof. Brian D.O. Anderson, Prof. Michel Gevers, Dr Venkatappa Sreeram, Prof. Sang Woo Kim and Prof. Yutaka Yamamoto. However, the majority of the work, approximately 70%, is my own.

A Madievski

February 1, 1995
Anton Madievski

Department of Systems Engineering
Research School of Information Sciences
and Engineering
The Australian National University
Canberra ACT 0200
Australia

Acknowledgements

Firstly, I thank Prof. Brian Anderson for the supervision and support he has provided during my time here in the Department of Systems Engineering. I am deeply grateful to Brian for introducing me to that mysterious science called Control and for showing great tolerance and patience while guiding me into a number of interesting areas of research, often in spite of my apparent efforts to resist him.

I would also like to thank Prof. Michel Gevers of *Université Catholique de Louvain*, Louvain la Neuve, Belgium, for his valuable expert advice and for fruitful collaboration during his visits to *The Australian National University* and his hospitality during my visit to Louvain la Neuve.

It is a pleasure to thank Dr Venkatappa Sreeram of *The University of Western Australia (Dubble Yeway)* for very enjoyable collaboration. I have learned a lot from him.

During my studies I have been fortunate to visit Université Catholique de Louvain, Belgium (Centre for Systems Engineering and Applied Mechanics), Delft University of Technology, the Netherlands (Mechanical Engineering Systems and Control Group), Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble, France (Laboratoire d'Automatique de Grenoble), Moscow State University, Russia (Department of Applied Mathematics) and BHP Research (Melbourne Laboratories). I would like to thank everyone concerned with these visits for their enthusiasm and support.

I also express my deep appreciation to the Australian National University and the Cooperative Research Centre for Robust and Adaptive Systems for their financial support.

For providing a very stimulating, exciting and enjoyable working environment, I extend my thanks to all the staff and students of Systems Engineering of the last few years. Special thanks to Mr James Ashton for his unrepeated ability to answer all my questions regarding the computer network.

I would also like to thank my parents and my grandmother who contributed to the research by providing an unmatched home environment and helping me to put in long office hours towards the thesis.

Lastly, I want to give thanks to all those I have called friend in the last few years. (This should not be considered in opposition to anyone mentioned above, but rather extends the list of people I feel very grateful to).

Abstract

In this thesis we are concerned with various problems of controller implementation, primarily concentrating our interest on sampled–data control systems and on controller order reduction.

We begin our investigation by presenting a new approach to establish a frequency–domain paradigm for the sampled–data control systems. The key idea is to associate a sampled–data system with a discrete system which is obtained from the original sampled–data system by very fast sampling at a multiple of the sampling frequency, followed by "blocking" or "lifting" to obtain a single–rate system. The approach is related to current approaches of Yamamoto and of Araki, Hagiwara and Ito.

The problem of an optimal finite wordlength state–space realization of a digital controller is investigated next. The closed loop to be considered consists of a continuous–time plant, a discrete–time controller, a sampler, a zero–order hold and an antialiasing filter. An effective algorithm is proposed to find the optimal sampled–data controller realization minimizing the sensitivity of the closed–loop performance with respect to coefficient errors in the state variable matrices of the controller realization. In order to get a tractable problem the fast sampling followed by lifting procedure needs to be applied. The procedure allows consideration of the system's inter–sample behaviour.

We proceed to investigation of the problem of controller order reduction aimed at preserving the closed–loop performance of a sampled–data closed–loop system. The fast sampling and lifting procedure allows capturing of the system's inter–sample behaviour and yields a time–invariant single–rate system; this then permits standard order reduction ideas to be applied. Special weighting functions aimed at preserving the closed–loop transfer function are obtained and weighted balanced truncation is used to reduce the controller.

An error bound for transfer function order reduction is derived next, when frequency weighted balanced truncation is the order reduction method. The bound is valid for both one–sided (input or output) and two–sided weighted balancing

approximations with stable weights, which can otherwise be arbitrary. Examples are studied to demonstrate effectiveness of the error bound.

We continue with showing by an example that frequency weighted balanced reduction procedure when applied to a scalar stable transfer function with input and output weights can result in an unstable reduced order model. A variation on the method is then presented which is guaranteed to yield stable reduced order models even when both input and output weightings are included. The method is a generalisation of a recently suggested technique and can handle weighting transfer functions which are proper rather than only strictly proper. A frequency response error bound for the proposed technique is also derived which is applicable for proper (including strictly proper) weighting functions.

An algorithm for transfer function order reduction is presented in the final part of the thesis, which generalizes the balanced stochastic truncation algorithm to allow for input weighting. An example illustrates use of the algorithm to secure smaller dB error in selected frequency bands through the introduction of the weighting.

Contents

Chapter 1

Introduction

1.1 Sampled–Data Control Systems

1.1.1 An Introduction to Sampled–Data Systems

The control of physical systems with a digital computer is becoming more and more common. Aircraft autopilots, mass–transit vehicles, oil refineries, paper–making machines and countless electromechanical servomechanisms are among the many existing examples. Furthermore, many new digital control applications are being stimulated by microprocessor technology including control of various aspects of automobiles and household appliances. Among the advantages of digital logic for control are the increased flexibility of the control programs and the decision–making or logic capability of digital systems, which can be combined with the dynamic control function to meet other system requirements.

The sampled–data control problems studied in this thesis are for feedback closed–loop systems. A typical topology of such a sampled–data system is shown in Fig. 1.1. The process to be controlled is called the plant and may

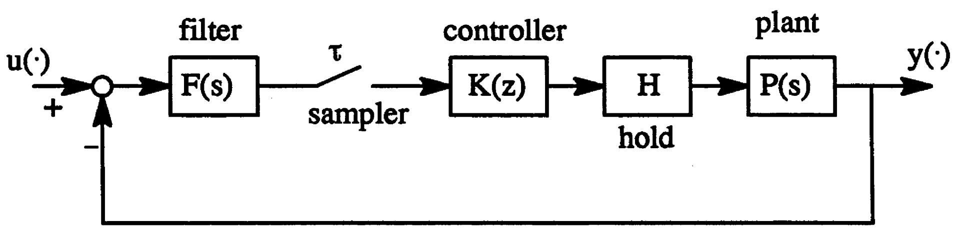


Figure 1.1: The sampled–data control system

be any of the physical processes mentioned above whose satisfactory response requires control action.

In this thesis we make the assumption that the sampling period (or sampling periods in the case of multi–rate sampling) are the same and fixed. This is in opposition to the *free running* scheme which assumes induced sampling upon completion of each cycle of the code execution.

In a real sense the problems of analysis and design of sampled–data controllers are concerned with taking account of the effects of the sampling time τ. If τ is extremely small (sampling frequency 50 or more times the system bandwidth), sampled–data systems are nearly continuous and continuous methods of analysis and design can be used. Many systems are originally conceived with fast sample rates in mind. However, as the design evolves, more demands are placed on the system, and the only way to accommodate the increased computer load is to slow down the sample rate. Furthermore, for cost–sensitive digital systems, the best design is the one with the lowest cost computer that will do the required job. That means the slowest speed computer. Slow sampling rate sampled–data systems cannot be treated as continuous–time systems and continuous methods of analysis and design are not applicable to the sampled–data systems.

Sampled–data systems cannot be treated as discrete–time systems either, for in the case of discretization of the sampled–data system the intersample behaviour of the system is ignored. We are interested in continuous–time performance of the plant and good behaviour of the plant at sampling instants does not guarantee satisfactory performance at the inter–sample periods.

Let us summarize to the point. A sampled–data system is a system having both discrete and continuous signals when a discrete–time controller is used to control a continuous–time plant. Interconnections of the controller and plant are made via a sampler with a fixed sampling time τ and a zero order hold keeping a discrete input value constant for the time τ. Naturally, while using the discrete–time controller, we are interested in continuous–time behaviour of the plant. That is why no standard continuous–time, nor discrete–time control analysis techniques are applicable to a sampled–data system. The need for sampled–data systems design and analysis special methods and theories becomes even more obvious when we think of the sampled–data systems as the most common and widespread control systems.

1.1.2 Frequency Response

The notion of *frequency response* plays a very important role in the theories of both continuous–time and discrete–time systems. That is why it is so important to define frequency response of a linear sampled–data system. The term frequency response is meant with the inter–sample behaviour of the sampled–data system taken into account, and therefore, the conventional approach of discrete–time approximation of the sampled–data system only at sampling instants cannot be applied.

A theory of sampled–data system frequency response based on such a definition will ideally allow for analysis and design of sampled–data controllers using standard concepts, theory and software.

A particular use of the transfer function or transfer function matrix of a linear time–invariant system is in computing the L_2 induced norm, which is related in a well–known way to the transfer function description. It turns out that the problem of computing the L_2 induced norm of a sampled–data system and related problems have been studied by many researchers, (e.g. Araki and Ito, 1992, 1993; Araki *et al.*, 1993; Bamieh and Pearson, 1992; Chen and Francis, 1990; Hara and Kabamba, 1990; Hayakawa *et al.*, 1992; Kabamba and Hara, 1990; Pearson *et al.*, 1991; Sivashankar and Khargonekar, 1992; Sun *et al.*, 1991; Tadmor, 1992; Toivonen, 1992; Yamamoto, 1993, 1994; Yamamoto and Araki, 1994; Yamamoto and Khargonekar, 1993), mainly using the technique of lifting an element in $L_2[0,\infty)$ up to that in $l_2^{L2[0,\tau]}$, where τ is the sampling period. In other words, many of these methods look at the time domain transfer characteristics between the input and output signals. In view of this, it might be appropriate to call them time domain approaches, or even $L_2[0,\tau]$–based approaches. The problem of computing the L_2 induced norm however is easier than defining a meaningful and practically useful – the two concepts are not the same – frequency response.

The major problem in defining a frequency response for a sampled–data system is that unlike continuous–time systems where a fixed frequency sinusoidal input produces a sinusoidal output at the same frequency, the sampled–data system output will not be a sine wave. The first approach undertaken to describe the sampled–data system frequency–domain behaviour was probably the modified z–transform due to Jury (Jury, 1964). Recently, there have been new attempts to develop a frequency–domain theory for sampled–data systems, e.g. Goodwin and Salgado (1992). Their approach was developed by Araki, Hagiwara and Ito (Araki and Ito, 1992, 1993, 1994; Araki *et al.*, 1993, 1994; Hagiwara *et al.*, 1993; Ito *et al.*, 1993), where a modified form of Fourier analysis was used.

Another approach described in Yamamoto (1993) and Yamamoto and Khargonekar (1993), is based upon the lifting technique (Friedland, 1960), which allows viewing of the sampled–data system as a time–invariant system by extending the input/output spaces to function spaces.

In Yamamoto and Araki (1994) it was proven that the notions of frequency response defined by the two methods are identical, at least for a magnitude point of view.

In this thesis we establish another approach to sampled–data system frequency response which takes inter–sample behaviour of the system into account. Unlike all other approaches based on continuous–time considerations, our approach is essentially discrete–time, but, nevertheless, describes sampled–data systems completely, but with a level of approximation which can be made arbitrarily small.

We base the sampled–data system frequency response definition on the idea of *fast sampling and lifting*. Fast sampling of continuous–time parts of the system N times faster than the sampling frequency τ^{-1} converts the sampled–data system to a discrete–time system which, nevertheless captures information on intersample behaviour of the original sampled–data system. The resulting fast–sampled discrete–time system is a multi–rate system with the sampling rates τ^{-1} and $N\tau^{-1}$. To convert this system to a single–rate system a procedure called lifting is applied. In the nutshell lifting is the combining of groups of N subsequent system input and output values into vectors. Clearly, that increases input and output dimensions of the system N times, but converts the multi–rate system to a slow (τ^{-1}) single–rate system. The beauty of the scheme is that for linear systems the lifting does not increase the order of the system.

If fast sampling is infinitely fast, the system would be described completely and our approach is actually like that of Yamamoto (Yamamoto, 1993, Yamamoto and Khargonekar, 1993). But to obtain a good approximation, we do not have to sample continuous–time entries of the system very fast; in Anderson and Keller (1994) it is argued that 5 times faster then the sampling frequency is usually enough.

The fast–sampling and lifting approach gives us the tool to handle sampled–data systems and armed with the method we may address a great many applied problems.

1.1.3 Finite Wordlength Sensitivity Minimization

One of the very important sampled–data control problems is the problem of choosing a realization of a digital compensator of a known transfer function, which ensures that the errors introduced into a sampled–data closed loop by using finite wordlength arithmetic in the compensator operation are minimized.

It is well known that a desired controller's transfer function can be implemented by any one of an infinite set of realizations of the controller. Though all these realizations are in principle equivalent since they yield the same transfer function, they have different numerical properties due to finite word length effects when they are implemented by a digital device (say, computer). Such factors as sensitivity and error propagation strongly affect closed–loop performance and are responsible for differences between desired ideal closed–loop characteristics and those actually obtained. A problem of great importance is to find the realization of the controller which achieves the best performance of the closed–loop system, i.e. gives the best approximation of the ideal closed loop behaviour.

Results on optimal realizations of filters (or "open–loop systems") minimizing some measure of performance degradation due to FWL errors date back to the late seventies. The first results were on realizations that minimise roundoff error propagation (Hwang, 1977, Mullis and Roberts, 1976). Realizations minimizing some measure of the transfer function sensitivity to coefficient errors took much longer to emerge (Thiele, 1986).

It was not until the late eighties that the problem of optimal controller realization minimizing closed loop performance degradation due to numerical errors was addressed. Solutions were proposed first for specific control schemes (LQG, pole placement), and more recently for general two degree of freedom controllers (Li and Gevers, 1990a, 1990b, 1991, Liu and Skelton, 1990, Liu *et al.*, 1992, Williamson and Kadiman, 1989). The last three references provide an optimal FWL LQG design (which includes an optimal realization in the design process). (Liu and Skelton (1990) and Liu *et al.* (1992) provide an optimal approach and Williamson and Kadiman (1989) provide a sub-optimal approach).

A survey of these results can be found in Gevers and Li (1993). The methods essentially differ in the choice of performance measure (either roundoff error propagation or transfer function sensitivity) and in the norms used to evaluate this performance degradation. In Gevers and Li (1993) a synthetic measure of performance degradation of a closed loop system, incorporating both roundoff errors and coefficient errors, was minimised with respect to all compensator realizations. The closed loop sensitivity minimization results in the above references all pertain to sensitivity measures of the closed loop **transfer function** with respect to controller parameter errors. In Li and Gevers (1993), a weighted sensitivity measure of the **closed loop poles** with respect to controller parameter errors is minimised.

The common feature of all these optimal controller realization results is that the system to be controlled is assumed to be described by a discrete time transfer function H(z). In most practical applications, a digital controller is used to control a continuous–time plant, using both a sampler and a hold device.

Any optimization using solely a discrete–time transfer function of the closed loop neglects the intersample system behaviour and particularly intersample ripple. The novel contribution presented in this thesis is to pose and solve a discrete time compensator realization problem for a continuous–discrete closed–loop system, in which the digital controller acts on the continuous time plant via a zero order hold device, and in which the tracking error of the continuous system is passed through an antialiasing filter and then sampled. With this continuous–discrete set up, the performance measure involves, of necessity, a hybrid operator: it is a measure of the sensitivity of the closed loop input–output operator to the parameters of the compensator realization.

In this thesis we establish the definitions of sensitivity "functions" (operators) and L_2 sensitivity measure of a closed loop system. Subsequently we study the finite–wordlength–optimal realization minimizing a measure of the sensitivity of the closed–loop operation with respect to controller coefficient errors. (No claim is made about FWL roundoff noise effects). The existence and uniqueness of an optimal solution are established. A recursive algorithm for obtaining the optimal solution is given. The fast–sampling and lifting procedure, which allows consideration of intersample behaviour of a closed–loop system, is applied to solve the problem.

1.2 Controller Order Reduction

The great importance and usefulness of controller reduction is now widely recognised and a great deal of attention has been paid to the subject over the past years. The main reason is that the LQG and $H\infty$ design procedures lead to controllers which have order equal to, or roughly equal to, the order of the plant (Anderson and Moore, 1971 for LQG). Often, controllers of lower order will result in acceptable performance, and will be desired for their greater simplicity.

1.2.1 Sampled–Data Controller Order Reduction

Model reduction by means of balanced realizations and Hankel–norm approximations has been studied by Moore (1981) and Glover (1984). Enns (1984a) introduced frequency weighting to balanced realizations and applied this approach for maintaining closed–loop stability (Enns, 1994b). Latham and Anderson (1985) have developed a frequency–weighted Hankel–norm technique for controller reduction. None of this earlier work explicitly treated sampled–data systems.

In this thesis, our objective is to apply a balanced realization controller order reduction method to sampled–data closed–loop systems to preserve the closed–loop behaviour.

A sampled–data control system is a periodically time–varying system. To replace this system by a time–invariant one *capturing intersample behaviour of the system*, one can apply the fast–sampling and lifting procedure.

There exist frequency–dependent weighting functions on the error between the original and reduced order controller transfer function matrices with the property that minimizing the weighted error corresponds approximately to minimizing an error between the two closed-loop transfer function matrices. We shall apply a weighted balanced realization technique to reduce the controller.

Unfortunately, reduction based on weighted balanced truncation is limited to (open-loop) stable controllers. One way to handle the problem in the unstable case is to additively decompose the full order controller transfer function into stable and completely unstable parts with the balanced realization technique applied to the stable part only. The system so obtained is a time-invariant single-rate discrete-time system, (with sample interval equal to that of the controller).

1.2.2 Frequency-Weighted Balanced Truncation

Controller design methods for physical systems with high order models normally result in a high order controller and, for many reasons, it is desirable to reduce the controller order, i.e. to find a controller of a lower order, performing satisfactorily in certain sense. Examples of these performance criteria include (but are not limited to) preserving of the closed-loop stability robustness and closed-loop transfer function (Anderson and Liu, 1989). All the controller reduction problems aimed at achieving these goals can be stated as problems of frequency weighted transfer function order reduction with the frequency weighting implying that it is important for the reduced order controller to approximate the original full order controller better at some frequencies, than at others.

Enns (1984a) has presented a scheme for reducing a stable high order model with frequency weighting, based on a modification of balanced truncation (Moore, 1981). The method, known as frequency weighted balanced truncation, may use input weighting, output weighting, or both. With only one weighting present, stability of the reduced order model is guaranteed. With both weightings present, there is no proof of reduced order model stability, although no example of instability has been reported so far. To overcome the potential drawback of instability, Lin and Chiu (1992) proposed a new frequency weighted balanced reduction technique. They showed that the reduced-order models obtained by their technique are necessarily stable when both input and output weightings are included. However, in the process of proving stability, they made two assumptions: (i) the input and output weighting functions are strictly proper, (ii) the input weighting realization is in input balanced form and output weighting realization is in output balanced form. Although, the second assumption does not affect the generality of their technique, the first assumption does. Furthermore, in controller reduction applications (Anderson and Liu, 1989, Kim *et al.*, 1994), usually the weighting functions are proper and not strictly proper.

In the thesis, we show by examples that Enns' technique may give an unstable reduced order model or may not give any reduced order model of a particular order. We then propose a new frequency weighted balanced truncation technique which is guaranteed to yield stable reduced order models even when both input and output weightings are included. The proposed technique is essentially a simple generalisation of Lin and Chiu's technique and can handle weighting transfer functions which are proper.

1.2.3 Error Bound for Reduction Using Frequency–Weighted Balanced Truncation

It is highly desirable to be able to predict an error in approximating an original full order controller by a reduced one, or at least to know an error bound. If such a prior error bound is known, it allows an intelligent estimation of the effect of a given degree reduction. For example, given a maximum acceptable error value, one may be able to determine a minimum order the controller can be reduced to, or, given a desired order of a reduced controller, one can predict an error this reduction brings (or at least bound it). This makes trade–off between the order of a reduced controller and the related approximation error easier to deal with.

Furthermore, knowledge of a prior error formula might allow one to compare alternative approaches to controller reduction, employing different frequency weightings aimed at achieving different objectives. For example, it allows one to optimise a trade–off between a stability margin achieved and the closed–loop transfer function being preserved, both possible objectives of the order reduction process (Anderson and Liu, 1989).

Lower and, more importantly, upper frequency domain error bounds for the balanced truncation approximation in the non–weighted case are well known and have been described in (Enns, 1984 a&b, Glover, 1984). However, no error bound formula has been available for the balanced truncation frequency–weighted problem.

One of the contributions of this thesis is that an upper error bound for frequency weighted balanced controller reduction is obtained. The bound is valid for both one–sided (input or output) and two–sided weighted balancing approximation for stable weights which can otherwise be arbitrary.

Furthermore, we also present frequency response error bounds for the proposed generalised Lin and Chiu's technique.

1.2.4 Multiplicative Frequency–Weighted Order Reduction

In this thesis, we consider a class of frequency weighted model reduction problems which finds an r–th order transfer function $V_r(s)$ aimed at minimizing the following error index:

$$E_W = \|V^{-1}[V-Vr]W\|_\infty \tag{1.2.1}$$

Here V(s) is a given stable transfer function of order greater than r, and W(s) is a given stable weighting function. With W(s) the identity, balanced stochastic truncation (BST) (Desai and Pal, 1984, Green, 1988a&b, Green and Anderson, 1990) is one method that can be used to find a $V_r(s)$ which is approximately minimizing. Our scheme with non–constant W(s) generalizes BST and we term it therefore weighted BST. We also term E_W of (1.2.1) the weighted multiplicative error.

The balanced stochastic truncation (BST) method which finds a reduced order transfer function minimizing or approximately minimizing a multiplicative error E_W with W(s)=1 was initiated by Desai and Pal (1984) and generalized by Green (1988a&b) and Green and Anderson (1990). The unweighted error index, involving so–called multiplicative approximation, is to be contrasted with the index $\|V-Vr\|_\infty$, which involves additive approximation. The former tends to produce an error which is flat in the dB sense (i.e. as a percentage) with frequency, while the latter tends to produce an absolute error with flat magnitude. Often a multiplicative approximation is preferred to the additive approximation, and a great many magnitude specifications are given in decibels (Integrated Systems, 1991); $\pm$1dB correspond to a multiplicative error of about 2%. Specification involving phase shift can also be regarded as multiplicative error statements; an error of $\pm$0.5 radians of phase shift is like a 5% multiplicative error also. Multiplicative error approximation rather than additive error approximation is also important in reducing high order models of plants, prior to design of a controller.

The 'robustness theorem' of Safonov et al. (Safonov *et al.*, 1988, Safonov and Chiang, 1989) also provides a compelling case for the importance of multiplicative error reduction of a plant model in control system design. Use of an unweighted criterion causes the multiplicative error to be approximately uniformly small, including at frequencies well above the closed-loop cut-off frequency, where larger multiplicative error in the plant model could be tolerated with no risk to stability. The introducing of frequency weighting to reflect this fact then allows smaller error to be obtained in the pass-band (which is helpful) at the expense of large error at the stop-band (which can be tolerable). Actually, in digital filter design, where the use of a multiplicative error is in general logical in, say, approximating a high order FIR filter by a low order IIR filter, deep in the stop-band multiplicative error probably also has much reduced relevance. Again therefore, there is case for introducing weighting into the multiplicative criterion.

The problem which finds $V_r(s)$ minimizing the weighted multiplicative error E_W could also (at least in principle) be solved by the two-sided weighted balanced truncation method (Enns, 1984a) in which the input weighting is W(s) and the output weighting $V^{-1}(s)$. However the calculation of this approximation is more complicated than that of the weighted balanced stochastic truncation, because the degrees with which one is working are certainly higher. Furthermore this method is not applicable when V(s) is non-minimum phase because then the output weighting is unstable. Even when V(s) is minimum phase, so that $V^{-1}(s)$, V(s) and W(s) are all stable, the two-sided weighted balanced truncation method may yield an unstable $V_r(s)$, in contrast to the $V_r(s)$ weighting when the scheme of this thesis is used.

One of the contributions of this thesis is to develop an algorithm for the input-weighted balanced stochastic truncation method. This algorithm can be extended to the output-weighted and the two-sided weighted cases.

1.3 Outline of the Thesis

A brief outline of the contents of this thesis is as follows. We begin in Chapter 2 with establishing another approach to sampled-data system frequency response which takes inter-sample behaviour of the system into account. Unlike all other approaches based on continuous-time considerations, our approach is essentially discrete–time, but, nevertheless, describes sampled–data systems completely, but with a controllable level of approximation.

We base the sampled–data system frequency response definition on the idea of fast sampling of continuous–time parts of the system followed by lifting of the obtained multi–rate system to convert it to a single–rate system. We present and then study the fast–sampling and lifting approach applied to the sampled–data system. A number of interesting general theoretical results on the method are presented. There are also several examples studied and analyzed.

The problem of an optimal finite wordlength state–space realization of a digital controller is investigated in Chapter 3. We establish the definitions of sensitivity "functions" (operators) and L_2 sensitivity measure of a closed loop system. Subsequently, we study the finite–wordlength–optimal realization minimizing a measure of the sensitivity of the closed–loop operation with respect to controller coefficient errors. (No claim is made about FWL roundoff noise effects). The existence and uniqueness of an optimal solution are established. A recursive algorithm for obtaining the optimal solution is given. Two numerical examples to confirm theoretical results followed by some concluding remarks are presented.

The problem of controller order reduction aimed at preserving the closed–loop performance of a sampled–data closed–loop system is investigated in Chapter 4. A time–invariant system results from applying the fast sampling and lifting procedure to a sampled–data system. In this chapter we obtain the weighting functions for preserving the closed–loop transfer function and actually reduce the controller by the weighted balanced truncation method. A practical example to confirm the approach is presented, followed by some concluding remarks.

In Chapter 5 we derive an upper error bound for transfer function order reduction, when frequency weighted balanced truncation is the order reduction method. The bound is valid for both one–sided (input or output) and two–sided weighted balancing approximation for stable weights which can otherwise be arbitrary. After reviewing the algorithm for weighted balanced reduction we derive the error bound formula itself. An example (showing tightness of the bound) is given, followed by some concluding remarks.

In Chapter 6, we show by examples that frequency–weighted balanced truncation may give an unstable reduced order model or may not give any reduced order model of a particular order. We then propose a new frequency weighted balanced truncation technique which is guaranteed to yield stable reduced order models even when both input and output weightings are included. The proposed method is essentially a generalisation of Lin and Chiu's (1992) technique and can handle weighting transfer functions which are proper rather than strictly proper. Furthermore, a frequency response error bound for the proposed technique is also derived which is applicable to proper (including strictly proper) weighting functions. Several examples are presented to compare the Enn's technique and the new scheme.

The principal concern of Chapter 7 is developing of an algorithm for the input–weighted balanced stochastic truncation method. After reviewing the algorithm for non–weighted balanced stochastic truncation we present the algorithm for transfer function order reduction, which generalizes the balanced stochastic truncation algorithm to allow for input weighting. This algorithm can be extended to the output–weighted and the two–sided weighted cases. An example illustrates use of the algorithm to secure smaller dB error in selected frequency bands through the introduction of the weighting.

To end the thesis, some concluding remarks and possible further research directions are given in Chapter 8.

Most of the proofs are given in the appendices.

Basic research is what I am doing when
I don't know what I am doing.

Werner von Braun
in R.L.Weber *A Random Walk in Science*

Chapter 2

Frequency Response of Sampled–Data Systems

2.1 Introduction

A new approach to establish a frequency–domain paradigm for sampled–data control systems is presented in this Chapter. The key idea is to associate a sampled–data system with a discrete system which is obtained from the original sampled–data system by very fast sampling followed by lifting to convert the sampled multi–rate system to a single–rate one. The approach is related to current approaches of Yamamoto and of Araki, Hagiwara and Ito.

A theory of sampled–data system frequency response based on such an approach will ideally allow for analysis and design of sampled–data controllers using standard concepts, theory and software.

We base the sampled–data system frequency response definition on the idea of fast sampling of continuous–time parts of the system followed by lifting of the obtained multi–rate system to convert it to a single–rate system. If fast sampling is infinitely fast, the system would be described completely and our approach becomes that of Yamamoto (Yamamoto, 1993, Yamamoto and Khargonekar, 1993). But to obtain a good approximation, we do not have to sample continuous–time entries of the system very fast; in Anderson and Keller (1994) it is argued that 5 times faster than the sampling frequency is usually enough.

2.2 The Operations of Fast Sampling and Lifting

To make the definition of the sampled–data system frequency response we use a two–step procedure. Our aim now is to introduce the procedure, which involves fast sampling and lifting operations on a sampled–data system. The effect is to replace the periodically time–varying system (with continuous–time input and output) by a time–invariant discrete–time system.

The end–result is a discrete–time approximation of the system. The fast sampling interval τ/N is chosen to be a submultiple N of the system sampling time τ, and the fast–sampled system is a multi–rate N–periodic discrete–time system. Lifting involves passing from an N–periodic linear $p \times m$ discrete–time sampled system, to an equivalent $pN \times mN$ discrete–time linear time–invariant system; the equivalence is an isomorphism of the unlifted and lifted systems in the sense that a number of essential algebraic and analytic properties of the systems are preserved. In particular, the lifted system is stable if and only if the N–periodic system is stable, and in this case certain operator norms (including those associated with regarding the system as an operator mapping square–summable inputs to square–summable outputs) are equal.

Normally, for computational purposes τ/N is chosen to be smaller than the fastest significant time constant of the sampled–data system, e.g. the inverse of $20 \times$ closed–loop bandwidth.

General system description

Let us consider a sampled–data control system shown in Fig. 2.1, where G(s) is a continuous–time plant and K(z) is a discrete–time controller. The controller input is sampled with the sampling time τ and its discrete–time output is converted to a continuous–time signal by using a zero–order hold H(s).

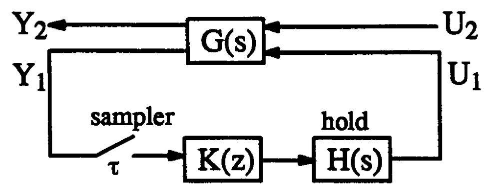


Figure 2.1. Sampled–data control system

The plant G(s) maps m_1– and m_2–element input vector–functions of time $U_1(\cdot)$ and $U_2(\cdot)$ into p_1– and p_2–element output vector–functions of time $Y_1(\cdot)$ and $Y_2(\cdot)$. In some works a filter with stable, no direct feedthrough transfer function is used on Y_1 to supress the aliasing effect of the sampler. Here we assume an anti–aliasing filter is absorbed into the plant.

Introduction of fast sampling of a continuous–time sub–system

Consider the $m(=m_1+m_2)$–input, $p(=p_1+p_2)$–output continuous–time sub–system G of the sampled–data system. Given state–space realizations of the sub–system as

$$G(s)=C(sI-A)^{-1}B+D, \tag{2.2.1}$$

where

$B = [B_1 \quad B_2]$, $C = [C_1^T \quad C_2^T]^T$ and $D = \begin{bmatrix} 0 & 0 \\ D_{21} & D_{22} \end{bmatrix}$

(the zeros in D are due to a no direct feedthrough anti–aliasing filter being absorbed into G). The state–space realization of the τ/N–sampled version of G is

$$g(z_N)=C(z_NI-a)^{-1}b+D, \tag{2.2.2}$$

where

$$a = \exp\,(A\ \tau/N), \tag{2.2.3}$$

$$b=\int_0^{\tau/N}\exp(A\ t)dt\ B. \tag{2.2.4}$$

The independent variable z_N is used to emphasize the τ/N spacing of the associated time sequence.

Equivalently, g (the τ/N-sampled version of G) is obtained as shown in Fig. 2.2 by introducing a zero-order hold at length τ/N and sampler at rate τ/N.

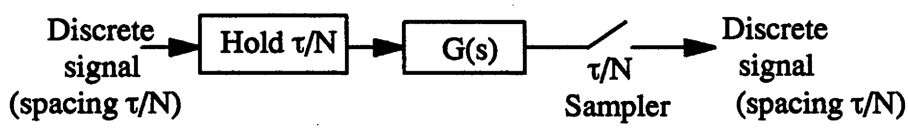


Figure 2.2. Fast-sampling of G

In rough terms, the fast-sampled sub-system g resembles the continuous-time sub-system G not only at the τ-distant sampling instants, but also at N−1 (τ/N)-equidistant points between every pair of subsequent τ-distant sampling moments. Hence, the fast-sampled sub-system has the potential to capture the inter-sample behaviour of the original sub-system. (Since G is a sub-system of a sampled-data system, we can apply the term inter-sample behaviour to G meaning synchronization with the τ-rate sampler).

The fast-sampled system g has a sampling rate N times faster than that of the original sampled-data system. To convert the fast-sampled system to the sampling rate of the sampled-data system (to "slow" rate), lifting needs to be applied.

Lifting of fast-sampled continuous-time sub-system

The lifting procedure is in fact simply a re-organization of the input and output values of the system, such that N subsequent input/output values are re-organized into an input/output vector.[1] This way the input/output vectors arrive/excite N times less frequently (every τ, not τ/N seconds), but no input/output value is lost. Clearly, this procedure increases the input/output dimensions N times, but the order of the system remains the same.

More precisely, if the input signals to $g(z_N)$ are

$$\begin{bmatrix} U_1[0] \\ U_2[0] \end{bmatrix}, \quad \begin{bmatrix} U_1[\tau/N] \\ U_2[\tau/N] \end{bmatrix}, \ldots \quad \begin{bmatrix} U_1[k\tau/N] \\ U_2[k\tau/N] \end{bmatrix}, \ldots$$

1. We will introduce formally the stacking (and unstacking) operators later in sub-section 2.3.1

the input signals to the lifted version are

$$\begin{bmatrix} \tilde{U}_1[0] \\ \tilde{U}_2[0] \end{bmatrix} = \begin{bmatrix} U_1[0] \\ \vdots \\ U_1[(N-1)\tau/N] \\ U_2[0] \\ \vdots \\ U_2[(N-1)\tau/N] \end{bmatrix}, \quad \begin{bmatrix} \tilde{U}_1[\tau] \\ \tilde{U}_2[\tau] \end{bmatrix} = \begin{bmatrix} U_1[\tau] \\ \vdots \\ U_1[\tau+(N-1)\tau/N] \\ U_2[\tau] \\ \vdots \\ U_2[\tau+(N-1)\tau/N] \end{bmatrix}, \ldots,$$

$$\begin{bmatrix} \tilde{U}_1[q\tau] \\ \tilde{U}_2[q\tau] \end{bmatrix} = \begin{bmatrix} U_1[q\tau] \\ \vdots \\ U_1[q\tau+(N-1)\tau/N] \\ U_2[q\tau] \\ \vdots \\ U_2[q\tau+(N-1)\tau/N] \end{bmatrix}, \ldots$$

Similarly, if the output signals from $g(z_N)$ are

$$\begin{bmatrix} Y_1[0] \\ Y_2[0] \end{bmatrix}, \quad \begin{bmatrix} Y_1[\tau/N] \\ Y_2[\tau/N] \end{bmatrix}, \ldots \quad \begin{bmatrix} Y_1[k\tau/N] \\ Y_2[k\tau/N] \end{bmatrix}, \ldots$$

the output signals of the lifted version $\mathcal{G}(z)$ are

$$\begin{bmatrix} \tilde{Y}_1[0] \\ \tilde{Y}_2[0] \end{bmatrix} = \begin{bmatrix} Y_1[0] \\ \vdots \\ Y_1[(N-1)\tau/N] \\ Y_2[0] \\ \vdots \\ Y_2[(N-1)\tau/N] \end{bmatrix}, \quad \begin{bmatrix} \tilde{Y}_1[\tau] \\ \tilde{Y}_2[\tau] \end{bmatrix} = \begin{bmatrix} Y_1[\tau] \\ \vdots \\ Y_1[\tau+(N-1)\tau/N] \\ Y_2[\tau] \\ \vdots \\ Y_2[\tau+(N-1)\tau/N] \end{bmatrix}, \ldots,$$

$$\begin{bmatrix} \tilde{Y}_1[q\tau] \\ \tilde{Y}_2[q\tau] \end{bmatrix} = \begin{bmatrix} Y_1[q\tau] \\ \cdot \\ \cdot \\ \cdot \\ Y_1[q\tau+(N-1)\tau/N] \\ Y_2[q\tau] \\ \cdot \\ \cdot \\ \cdot \\ Y_2[q\tau+(N-1)\tau/N] \end{bmatrix}, \ldots$$

Fig. 2.3 depicts the relations between $g(z_N)$ and $\mathcal{G}(z)$, the lifted version.

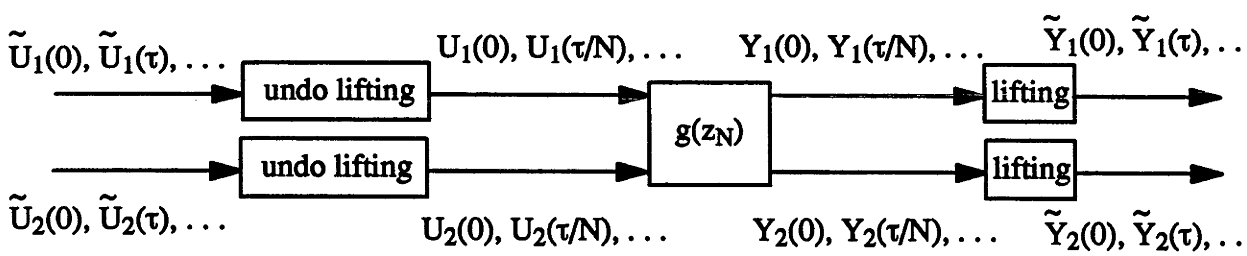


Figure 2.3. Lifting of g

The state-space realization of the mN-input, pN-output discrete-time lifted sub-system $\mathcal{G}$ can be written in the form

$$\mathcal{G}(z)=\mathcal{C}(zI-\mathcal{A})^{-1}\,\mathcal{B} + \mathcal{D} \tag{2.2.5}$$

where

$$\mathcal{A}= a^N, \tag{2.2.6}$$

$$\mathcal{B}=[a^{N-1}b \quad \ldots \quad ab \quad b], \tag{2.2.7}$$

$$\mathcal{C}=[C^T \quad a^T C^T \quad \ldots \quad (a^T)^{N-1} C^T]^T, \tag{2.2.8}$$

$$\mathcal{D} = \begin{pmatrix} D & 0 & \dots & 0 \\ C\,b & D & \dots & 0 \\ \dots & \dots & \dots & 0 \\ Ca^{N-2}b & Ca^{N-3}b & \dots & D \end{pmatrix} \tag{2.2.9}$$

The variable z is used for $\mathcal{G}(z)$ since the associated time sequences have spacing τ.

The realization ($\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$) of $\mathcal{G}(z)$ is minimal if and only if (a, b, C, D) is a minimal realization of $g(z_N)$. In turn, the realization (a, b, C, D) is minimal for almost all τ if and only if (A, B, C, D) is minimal. (e.g. Åström and Wittenmark, 1990).

Connection of a discrete-time controller with a sampler and hold to lifted system

We have described how fast sampling and lifting can be applied to G(s). Now we indicate how K(z) can be compatibly lifted. In particular, we shall argue that K(z) should be replaced by a time-invariant discrete-time $\mathcal{K}(z)$, with underlying sampling time τ, and with different input and output dimensions to K(z), so that the interconnection of Fig. 2.1 is replaced by the interconnection of Fig. 2.4, with $\mathcal{K}(z)$ simply obtainable in a way we now describe.

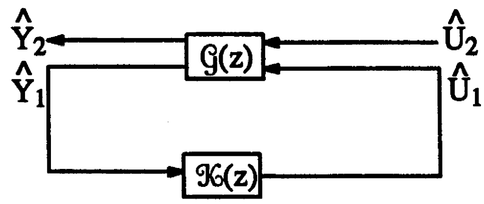


Figure 2.4. Fast-sampled and lifted control system

In Fig. 2.1, the input signal for K(z) is obtained by sampling the output Y_1 of G(s) every τ seconds. Such samples are contained in the output of $\mathcal{G}(z)$. More precisely, the output vector at time $q\tau$ of $\mathcal{G}(z)$ is

$$\begin{bmatrix} Y_1[q\tau] \\ Y_1[(qN+1)\tau/N] \\ \cdot \\ \cdot \\ \cdot \\ Y_1[(qN+N-1)\tau/N] \end{bmatrix}$$

and the sampled output Y_1 of G(s) is obtained by multiplying this vector by

$$(I_{m1}\ \ 0_{m1}\ \ 0_{m1}\ \ \ldots\ \ 0_{m1}) \in \mathbb{R}_{m1 \times m1N}, \qquad (2.2.10a)$$

Also, the output of K(z) is a constant over intervals of length τ. So the feedback input vector to $\mathcal{G}(z)$ at time $q\tau$, corresponding to inputs to $g(z_N)$ at times $q\tau$, $(qN+1)\tau/N, \ldots, (qN+N-1)\tau/N$, must contain equal entries; thus the input vector is given by

$$\begin{bmatrix} I_{p1} \\ I_{p1} \\ \cdot \\ \cdot \\ \cdot \\ I_{p1} \end{bmatrix} \times \text{ output of } K(z) \text{ at time } q\tau \qquad (2.2.10b)$$

All this means that the feedback input to $\mathcal{G}(z)$ and the output of $\mathcal{G}(z)$ used for generating the feedback signal are related by

$$\begin{bmatrix} U_1[q\tau] \\ U_1[(qN+1)\tau/N] \\ \cdot \\ \cdot \\ \cdot \\ U_1[(qN+N-1)\tau/N] \end{bmatrix} = \begin{bmatrix} I_{p1} \\ I_{p1} \\ \cdot \\ \cdot \\ \cdot \\ I_{p1} \end{bmatrix} K(z)\,[I_{m1}\, 0_{m1}\, 0_{m1} \ldots 0_{m1}] \begin{bmatrix} Y_1[q\tau] \\ Y_1[(qN+1)\tau/N] \\ \cdot \\ \cdot \\ \cdot \\ Y_1[(qN+N-1)\tau/N] \end{bmatrix}$$

(2.2.11)

$$\text{i.e. } \mathcal{K}(z) = \begin{bmatrix} I_{p1} \\ I_{p1} \\ \cdot \\ \cdot \\ \cdot \\ I_{p1} \end{bmatrix} K(z)\,[I_{m1}\, 0_{m1}\, 0_{m1} \ldots 0_{m1}]. \qquad (2.2.12)$$

Observe that if G(s) and K(z) are finite dimensional then $\mathcal{G}(z)$ and $\mathcal{K}(z)$ are finite dimensional too and therefore, so is the system of Fig. 2.4. Also, note that if K(z) stabilizes the sampled–data system of Fig. 2.1, then the system of Fig. 2.4 is closed–loop stable.

Summary of procedure for fast sampling and lifting of sampled-data system

Step 1: Compute fast–sampled version $g(z_N)$ of continuous–time G(s) according to (2.2.1–4).

Step 2: Compute discrete τ–spaced lifted version $\mathcal{G}(z)$ of discrete τ/N–spaced $g(z_N)$ according to (2.2.5–9).

Step 3: Compute a lifted signal sampler and a hold with lifted output and absorb them with discrete τ–spaced controller K(z) into a discrete τ–spaced $\mathcal{K}(z)$ as in (2.2.12).

Step 4: Replace G(s) by $\mathcal{G}(z)$ and K(z) together with the sampler and hold by $\mathcal{K}(z)$. The obtained discrete τ–spaced in time control system is illustrated in Fig. 2.4.

As the result of the substituting subsystems $\mathcal{K}(z)$ and $\mathcal{G}(z)$, the original periodically time–varying sampled–data system has been replaced (approximately) by a time–invariant system.

An input-output view of fast-sampling and lifting

Above, we have described the construction of a fast–sampled and lifted system in terms of applying these operations to subsystems of a closed–loop system, and using state–variable descriptions of the subsystems. We can also capture aspects of the procedure by considering fast–sampling and lifting applied to an arbitrary continuous–time, strictly proper, periodic system, see Fig. 2.5, with an impulse response description. The system, denoted by T, may be a sampled–data system, but this is not critical.

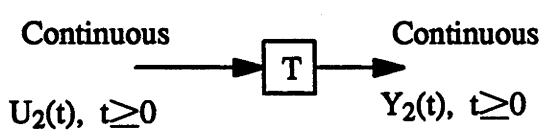


Figure 2.5. Strictly proper, periodic system

Fast sampling of the system results in the system depicted in Fig. 2.6.

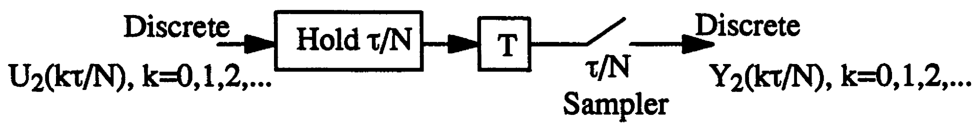


Figure 2.6. Fast-sampled system

Lifting of the obtained fast-sampled system which is simply re-arranging of the input and output sequences into vectors (or parallel instead of serial processing of inputs) delivers a system shown in Fig. 2.7, where i is the integer part of k/N.

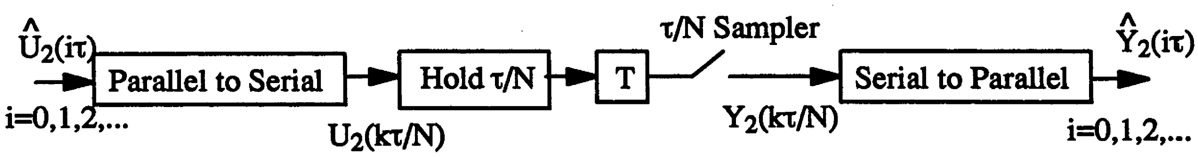


Figure 2.7. Lifted system

It is not hard to check that the scheme of Fig. 2.6 is periodically time-varying, with period τ, i.e. N times the underlying sampling interval. This means that the scheme of Fig. 2.7 has time-invariant input/output behaviour. Thus, the original periodically time-varying sampled-data system of Fig. 2.5 has been replaced by a time-invariant system of Fig. 2.8.

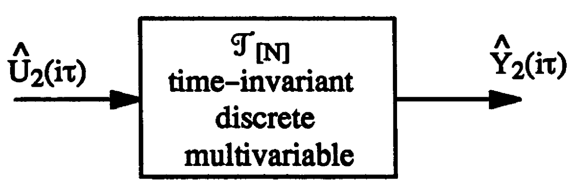


Figure 2.8. Fast-sampled and lifted system

If the impulse response associated with T is denoted by $T_c(t,s)$, then the impulse response of the fast-sampled system can be denoted by $T_d(k,l)$ where

$$T_d(k,l) = \int_{l\tau/N}^{(l+1)\tau/N} T_c(k\tau/N,s)ds \qquad k>l$$

[$T_d(k,l)$ is zero for $k \leq l$].

The impulse response associated with $\mathcal{T}_{[N]}$ is then

$$\mathcal{T}_{[N]}(k,l) = \begin{bmatrix} T_d(kN,lN) & T_d(kN,lN+1) & \cdots & T_d(kN,lN+N-1) \\ \vdots & \vdots & & \vdots \\ T_d(kN+N-1,lN) & T_d(kN+N-1,lN+1) & \cdots & T_d(kN+N-1,lN+N-1) \end{bmatrix} \tag{2.2.13}$$

(which is zero for $k<l$).

Because T is periodic with period τ, $T_c(t+\tau,s+\tau)=T_c(t,s)$ and $T_d(k+N,l+N)=T_d(k,l)$. It follows that $\mathcal{T}_{[N]}(k,l) = \mathcal{T}_{[N]}(k-l)$, displaying the time-invariance.

The discussion in this section is totally consistent with the earlier discussion where we worked with subsystems. The earlier discussion yields a state-variable description of $\mathcal{T}_{[N]}$, in contrast to the input-output discussion here.

2.3 Frequency Response of Fast-Sampled and Lifted Systems

In this section we study the fast-sampled/lifted system and establish its behaviourial similarity to the sampled-data system. We shall also show how the fast-sampled and lifted system operator can be used to (approximately) calculate the output of a sampled-data system to a given input.

2.3.1 Definitions and Background

We shall need the following definitions:

$\mathcal{S}^{\alpha}$: vector space of piecewise continuous functions from the time set $[0,\infty)$ to $\mathbb{R}^{\alpha}$ that are bounded on compact subsets of $[0,\infty)$ and continuous from the right at every point.

$\mathcal{S}^{\alpha}{}_{p}$: $L^{\alpha}{}_{p} \cap \mathcal{S}^{\alpha}$.

Hold of length τ/N
$H_{\tau/N} : l^{\alpha}_{p}(Z_+) \rightarrow L^{\alpha}_{p}[0,\infty)$. The p-summable sequence of α-vectors $\{v(0), v(\tau/N), v(2\tau/N), ...\}$ is mapped into the p-th power integrable $w(\cdot)$ with $w(t)=v(k\tau/N) \quad k\tau/N \leq t<(k+1)\tau/N$. Note that α is an arbitrary positive integer.

Sampler of interval τ/N
$S_{\tau/N} : \mathcal{S}^{\alpha}_{\infty} \rightarrow l^{\alpha}_{\infty}(Z_+)$. The bounded time function $w(\cdot)$, continuous from the right, is mapped into the sequence $\{w(0), w(\tau/N), w(2\tau/N), \ldots\}$ where, if $w(k\tau/N)$ is discontinuous, the value $\lim_{t \downarrow k\tau/N} w(t)$ is assumed.

Stacking operator
$St_{[N]} : l^{\alpha}_{p}(Z_+) \rightarrow l^{\alpha N}_{p}(Z_+)$. This operator lifts discrete-time α-vectors by stacking their values into αN vectors. Thus
$St_{[N]}\{v(0), v(\tau/N), v(2\tau/N), \ldots\} = \{\bar{v}(0), \bar{v}(\tau), \bar{v}(2\tau), \ldots\}$
where $\bar{v}(k\tau) = [v^T(k\tau) \quad v^T(k\tau+\tau/N) \quad \ldots \quad v^T(k\tau+(N-1)\tau/N)]^T$.

Unstacking operator
$Us_{[N]} : l^{\alpha N}_{p}(Z_+) \rightarrow l^{\alpha}_{p}(Z_+)$. This operator is the inverse of the stacking operator.

In the above definitions, there is an underlying time-scale associated with sequences $l^{\alpha}_{p}(Z_+)$ and $l^{\alpha N}_{p}(Z_+)$ etc. that has to be determined from the context.

It is evident that $S_{\tau/N} H_{\tau/N} : l^{\alpha}_{p}(Z_+) \rightarrow l^{\alpha}_{p}(Z_+)$ is the identity. However, $H_{\tau/N} S_{\tau/N}$ is not in general the identity operator, and indeed, $S_{\tau/N}$ may not map $L^{\alpha}_{p}[0,\infty)$ to $l^{\alpha}_{p}(Z_+)$ for $p<\infty$, see Chen and Francis (1989). Because our interest is particularly in square integrable/summable functions, we shall use the following important lemmata, which allow us to bypass this problem:

Lemma 2.1. Let T be an operator describing a causal linear system with impulse response $T(t,s)$, where $T(\cdot,\cdot)$ is continuous in $t \geq s$, and $\|T(t,s)\| \leq \alpha \exp[-\beta(t-s)]$ for some $\alpha>0$, $\beta>0$ and all $t \geq s$. Then

(i) T is a bounded map on $L_p(-\infty,\infty)$, $1 \leq p \leq \infty$

(ii) T maps $L_p(-\infty,\infty)$ to $C(-\infty,\infty)$, $1 \leq p \leq \infty$

(iii) With $L_{pav}(-\infty,\infty)$ denoting functions which are L_p on intervals of fixed length τ, and with norm $\sup_{T} [\int_{T}^{T+\tau} \|\cdot\|^p dt]^{1/p}$, T is a bounded map on L_{pav}, $1 \leq p \leq \infty$ and maps L_{pav} to $C(-\infty,\infty)$.

Proof. See Appendix A.

In the above Lemma, there is an equicontinuity with respect to $u(\cdot)$ of fixed norm [L_p or L_{pav}] but not uniform continuity (with respect to time). We can achieve this for periodic T.

Corollary 2.1. Adopt the hypothesis of Lemma 2.1 and assume further that $T(t,s) = T(t+\tau, s+\tau)$ for some τ. Suppose that $y(t) = \int_{-\infty}^{t} T(t,s)\, u(s)\, ds$ for $u(\cdot) \in L_p(-\infty,\infty)$. Then given $\varepsilon>0$, there exists $\bar{\delta}(\varepsilon)$ such that

$$\|y(t+\delta) - y(t)\| < \varepsilon \, \|u\|_p \qquad \forall \quad \delta < \bar{\delta}, \forall\, t.$$

Further, the same equicontinuity result holds if $u \in L_p(-\infty,\infty)$ is replaced by $u \in L_{pav}(-\infty,\infty)$.

Proof. See Appendix B.

Lemma 2.2. Let T be as in Lemma 2.1. Let H_h, S_h denote the hold of length h and sample of length h.

Then

(i) $H_h\, S_h\, T$ is a bounded operator on $L_p(-\infty,\infty)$, $1 \leq p \leq \infty$

(ii) $H_h\, S_h\, T$ is a bounded operator on $L_{pav}(-\infty,\infty)$, $1 \leq p \leq \infty$.

Proof. See Appendix C.

The last result we need is the follows:

Lemma 2.3. With the same hypothesis as Lemma 2.2,

$$\lim_{h \to 0} \|H_h\, S_h\, T - T\| = 0$$

in the $L_p(-\infty,\infty)$ induced norm and $L_{pav}(-\infty,\infty)$ induced norm for $1 \leq p \leq \infty$.

Proof. See Appendix D.

2.3.2 Sampled-Data and Fast-Sampled/Lifted Systems Norm Equivalence

In this sub-section we indicate convergence of the norm of the fast-sampled and lifted system to the norm of the sampled-data system. We also relate norms for time-invariant systems for fixed frequency.

It is intuitively clear that the performance of the fast-sampled/lifted system will in some way mimic that of the original sampled-data system. The following theorem is a partial extension of the results of Anderson and Keller, 1994. (An outline of the proof can be found in Appendix E).

Theorem 2.1. Let T be the operator describing the sampled-data system of Fig. 2.1 with the plant satisfying Condition E (from Appendix E) and with the closed-loop system stable, so that T maps $L_2^{m_2}[0,\infty)$ to $L_2^{p_2}[0,\infty)$. Let $\mathcal{T}_{[N]}$ be the operator describing the system of Fig. 2.4 obtained by fast-sampling with interval τ/N and lifting of the sampled-data system and mapping from $l_2^{m_2N}(Z_+)$ to $l_2^{p_2N}(Z_+)$. Then for all $p \in [1,\infty]$ the limit of the induced norm of $\mathcal{T}_{[N]}$ converges to the induced norm of T when the over-sampling coefficient $N \to \infty$:

$$\lim_{N\to\infty} \|\mathcal{T}_{[N]}\|_p = \|T\|_p$$

The next theorem establishes a close connection in behaviour of sampled-data and fast-sampled/lifted systems at a particular frequency. It shows that our approach is reasonable and sensible, for when it is applied to a continuous-time system it gives a meaningful result. Also, the theorem is of great importance for applications.

Theorem 2.2. Let G(s) be a SISO continuous-time system given in (2.2.1) and $\mathcal{G}_{[N]}(z)$ be its fast-sampled (with the sampling time τ/N) and N-lifted version given by (2.2.5). Then

$$\lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{G}_{[N]}^*(e^{j\omega\tau})\, \mathcal{G}_{[N]}(e^{j\omega\tau})\} = \max_{k=0,1,2,\ldots} |G(j(\omega+2\pi k/\tau))|.$$

Proof. Let $g(z_N)$ given by (2.2.2) denote the fast-sampled (with the sampling time τ/N) version of a continuous-time system G.

Then we claim first that $\mathcal{G}_{[N]}^*(e^{j\omega\tau})\ \mathcal{G}_{[N]}(e^{j\omega\tau})$ has N eigenvalues $g(e^{-j(2\pi k+\omega\tau)/N})g(e^{j(2\pi k+\omega\tau)/N})$ with the corresponding eigenvectors $(1 \quad e^{j(2\pi k+\omega\tau)/N} \quad e^{j2(2\pi k+\omega\tau)/N} \quad \dots \quad e^{j(N-1)(2\pi k+\omega\tau)/N})^T$, $k = 0, 1, \dots, N-1$.

To prove this statement, let $\{h_p\}$ denote the impulse response associated with $g(z_N)$, so that

$$g(e^{j\omega\tau/N}) = \sum_{p=1}^{\infty} h_p\, e^{-jp\omega\tau/N}.$$

The (n,m) entry of the impulse response of the lifted system is by (2.2.13) precisely h_{pN+n-m}, so that the (n,m)–th element of $\mathcal{G}_{[N]}(e^{j\omega\tau})$ can be written in terms of h_p as

$$[\mathcal{G}_{[N]}(e^{j\omega\tau})]_{n,m} = \sum_{p=1}^{\infty} h_{pN+n-m}\, e^{-jp\omega\tau}.$$

Multiplying $\mathcal{G}_{[N]}(e^{j\omega\tau})$ and the eigenvector $(1 \quad e^{j(2\pi k+\omega\tau)/N} \quad e^{j2(2\pi k+\omega\tau)/N} \quad \dots \quad e^{j(N-1)(2\pi k+\omega\tau)/N})^T$ and calculating the n–th element of the product–vector one can obtain:

$$\sum_{m=1}^{N} [\mathcal{G}_{[N]}(e^{j\omega\tau})]_{n,m}\, e^{j(m-1)(2\pi k+\omega\tau)/N}$$

$$= \sum_{m=1}^{N} \sum_{p=1}^{\infty} h_{pN+n-m}\, e^{j(-pN-n+m)(2\pi k+\omega\tau)/N}\, e^{j(n-1)(2\pi k+\omega\tau)/N}$$

$$= g(e^{j(2\pi k+\omega\tau)/N})\, e^{j(n-1)(2\pi k+\omega\tau)/N}$$

and that proves that $g(e^{j(2\pi k+\omega\tau)/N})$ is the corresponding eigenvalue.

A similar calculation shows that $\mathcal{G}_{[N]}^*(e^{j\omega\tau})$ has the same eigenvector, with associated eigenvalue $g(e^{-j(2\pi k+\omega\tau)/N})$. It then easily follows that $\mathcal{G}_{[N]}^*(e^{j\omega\tau})\ \mathcal{G}_{[N]}(e^{j\omega\tau})$ has the same eigenvector, with eigenvalue $g(e^{-j(2\pi k+\omega\tau)/N})g(e^{j(2\pi k+\omega\tau)/N})$.

On the other hand,

$$g(e^{-j(2\pi k+\omega\tau)/N})g(e^{j(2\pi k+\omega\tau)/N})$$
$$= D^2 + DC\{(e^{j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1} + (e^{-j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1}\} \int_0^{\tau/N} e^{At}dt\, B$$

$$+ C(e^{j(2\pi k+\omega\tau)/N} I - e^{A\tau/N})^{-1} \int_0^{\tau/N} e^{At/N}dt \; BC(e^{-j(2\pi k+\omega\tau)/N} I - e^{A\tau/N})^{-1} \int_0^{\tau/N} e^{At}dt \; B.$$

Performing simple calculations, one can obtain

$$\lim_{N\to\infty} \{(e^{j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1} + (e^{-j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1}\} \int_0^{\tau/N} e^{At}dt$$

$$= (j(\omega+2\pi k/\tau)I-A)^{-1} + (-j(\omega+2\pi k/\tau)I-A)^{-1}$$

and

$$\lim_{N\to\infty} (e^{j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1} \int_0^{\tau/N} e^{At}dt \; BC(e^{-j(2\pi k+\omega\tau)/N}I - e^{A\tau/N})^{-1} \int_0^{\tau/N} e^{At}dt$$

$$= (j(\omega+2\pi k/\tau)I-A)^{-1}BC(-j(\omega+2\pi k/\tau)I-A)^{-1}.$$

It follows that

$$\lim_{N\to\infty} g(e^{-j(2\pi k+\omega\tau)/N})g(e^{j(2\pi k+\omega\tau)/N}) = |G(j(\omega+2\pi k/\tau))|^2.$$

Thus, the limit of the maximum singular value $\lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{G}_{[N]}^*(e^{j\omega\tau})\, \mathcal{G}_{[N]}(e^{j\omega\tau})\}$ can be written as $\max_{k=0,1,2,\ldots} |G(j(\omega+2\pi k/\tau))|$ and that concludes the proof.

Remark 2.1. When D=0, i.e. $G(j\omega)$ has no direct feedthrough, the maximum $\max_{k=0,1,2,\ldots} |G(j(\omega+2\pi k/\tau))|$ is achieved for k=0 if the sampling time τ is chosen short enough to avoid the aliasing effect. Thus, for small enough τ we can write:

$$\lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{G}_{[N]}^*(e^{j\omega\tau})\, \mathcal{G}_{[N]}(e^{j\omega\tau})\} = |G(j\omega)|. \tag{2.3.1}$$

We are however interested in a more general result, dealing with convergence of operators and not just their norms. The form of this general result is motivated by two considerations:

(a) we are interested in examining the collection of operators $\mathcal{T}_{[N]}$ as $N\to\infty$; since the domain and range depend on N, it is convenient to work with operators closely related to the $\mathcal{T}_{[N]}$ which for all N have the same domain and range

(b) a key reason for introducing the $\mathcal{T}_{[N]}$ is that they purport to represent in some way the sampled–data system T. More precisely, we would hope that the calculation of the response of T to a prescribed continuous input could be somehow effected (approximately, with the approximation error decreasing to zero as $N \to \infty$) using $\mathcal{T}_{[N]}$. Since T however has a different domain and range from $\mathcal{T}_{[N]}$, it is clear that we need to modify $\mathcal{T}_{[N]}$ in some way for it to approximate T.

In the remainder of this section, our goal is to define a modification to $\mathcal{T}_{[N]}$ and prove a convergence result of a sequence of operators [all of which naturally have the same domain and range]. Following this, we shall indicate how the modification of $\mathcal{T}_{[N]}$ allows us to (approximately) calculate the output of T in response to a given input.

2.3.3 $\mathcal{T}_{[N]}$, the Modified $\mathcal{T}_{[N]}$ and their Relation to T

We now return to consider the schemes of Fig. 2.1 and Fig. 2.4, described by the periodically time–varying operator T and time–invariant operator $\mathcal{T}_{[N]}$. An examination of the way the discrete–time time–invariant operator $\mathcal{T}_{[N]}$ is constructed from T shows easily that

$$\mathcal{T}_{[N]} = St_{[N]}\ S_{\tau/N}\ T\ H_{\tau/N}\ Us_{[N]} \tag{2.3.2}$$

As already commented, the domain and range of the operators $\mathcal{T}_{[N]}$ are N–dependent. Let us define the following operator, obtained from $\mathcal{T}_{[N]}$ by operator pre–multiplication and post–multiplication with the multiplying elements designed to (as best as possible) cancel the $H_{\tau/N}\ Us_{[N]}$ post–multiplication of T and the $St_{[N]}\ S_{\tau/N}$ pre–multiplication of T, and at the same time designed to yield a common domain and range:

$$\tilde{\mathcal{T}}_{[N]}\ F = [H_{\tau/N}\ Us_{[N]}]\ \mathcal{T}_{[N]}\ [St_{[N]}\ S_{\tau/N}]\ F \tag{2.3.3}$$

Here F is an operator meeting the constraints of Corollary 2.1 with $\alpha = \beta = m_2$. It is immediately evident that with T and thus $\tilde{\mathcal{T}}_{[N]}$ stable, $\tilde{\mathcal{T}}_{[N]}$ F maps $L^{m_2}{}_p[0,\infty)$ to $L^{p_2}{}_p[0,\infty)$, $1 \leq p \leq \infty$. Thus the common domain and range are achieved.

Observe using (2.3.2) that

$$\begin{aligned}\tilde{\mathcal{T}}_{[N]}\,F &= H_{\tau/N}\,Us_{[N]}\,St_{[N]}\,S_{\tau/N}\,T\,H_{\tau/N}\,Us_{[N]}\,St_{[N]}\,S_{\tau/N}\,F \\ &= [H_{\tau/N}\,S_{\tau/N}]\,T\,[H_{\tau/N}\,S_{\tau/N}]\,F.\end{aligned}$$

We now have the following result:

Theorem 2.3. Let T and $\mathcal{T}_{[N]}$ be as described in Theorem 2.1, and suppose that T has no direct feedthrough. Let $\tilde{\mathcal{T}}_{[N]}\,F$ be constructed from $\mathcal{T}_{[N]}$ according to (2.3.3), where F is an operator satisfying the conditions of Lemma 2.1 - Corollary 2.1 with $\alpha = \beta = m_2$. Then

$$\lim_{N\to\infty} \| (T-\tilde{\mathcal{T}}_{[N]})F\|_p = 0$$

for $1 \leq p \leq \infty$.

Proof.

$$\begin{aligned}(T - \tilde{\mathcal{T}}_{[N]})F &= TF - (H_{\tau/N}\,S_{\tau/N})\,T\,(H_{\tau/N}\,S_{\tau/N})\,F \\ &= [TF - H_{\tau/N}\,S_{\tau/N}\,T\,F] + H_{\tau/N}\,S_{\tau/N}\,T\,[I - H_{\tau/N}\,S_{\tau/N}]\,F\end{aligned}$$

Now TF is an operator meeting the conditions of Lemma 2.1 and so the p-norm ($1 \leq p \leq \infty$) of the first summand approaches zero when $N \to \infty$ by Lemma 2.3. Also, $H_{\tau/N}\,S_{\tau/N}\,T$ is a bounded operator on the range-space of the operator $[I - H_{\tau/N}\,S_{\tau/N}]\,F$ acting on $L_p[0,\infty)$, $1 \leq p \leq \infty$, and $[I - H_{\tau/N}\,S_{\tau/N}]\,F$ has norm approaching zero when $N \to \infty$. Hence the norm of the second summand also approaches zero as $N \to \infty$. The triangle inequality then yields the result.

2.3.4 Approximate Calculation of System Response Using $\mathcal{T}_{[N]}$

The key to using the time-invariant operator $\mathcal{T}_{[N]}$ to calculate (with some approximation) the response of T to a given input lies in the result of Theorem 2.3.

A prerequisite is that the input to T must be bandlimited, (so that it can be regarded as being the output of a strictly causal stable F). Theorem 2.3 then says: T can be approximated by $\tilde{\mathcal{T}}_{[N]}$. In view of (2.3.3), this says that output of T can be computed (approximately) as follows:

- Sample the bandlimited m_2–dimensional input of T at rate τ/N
- Stack the sampled into m_2N–vectors, spaced τ apart
- Use this sequence as an input to the time–invariant system with description $\mathcal{T}_{[N]}$ and obtain the corresponding output sequence
- Unstack the output sequence
- Put the unstacked output sequence through an $H_{\tau/N}$ hold, to obtain a continuous–time signal

2.4 The Fast–Sampled Lifted System and its Relation to the System Description of Yamamoto

In this section we recall the frequency response description proposed by Yamamoto (Yamamoto, 1993, Yamamoto and Khargonekar, 1993) and establish behaviourial closeness of the system and proposed fast–sampled/lifted system.

As we will now show the operator $\tilde{\mathcal{T}}_{[N]}$ is closely related (through input and output spaces isomorphism) to the frequency response operator $\hat{\mathcal{T}}_Y$ of Yamamoto (Yamamoto, 1993, Yamamoto and Khargonekar, 1993). Notice that $\tilde{\mathcal{T}}_{[N]}$ is finite dimensional (in the input/output dimensions sense) while $\hat{\mathcal{T}}_Y$ is infinite dimensional; thus there is a real advantage in using $\tilde{\mathcal{T}}_{[N]}$.

2.4.1 Yamamoto's Approach

Let us recall briefly the frequency–response description due to Yamamoto. The approach is based on the idea of regarding input u(t) and output y(t) as the sequences of functions $u_k(\theta)$ and $y_k(\theta)$, $t \in [0,\infty)$, $k \in Z_+$, $\theta \in [0,\tau)$:

$$u_k(\theta) = u(k\tau+\theta)$$
$$y_k(\theta) = y(k\tau+\theta). \qquad (2.4.1)$$

One can regard the sequences $\{u_1(\cdot), u_2(\cdot), u_3(\cdot), \ldots\}$ and $\{y_1(\cdot), y_2(\cdot), y_3(\cdot), \ldots\}$ as lifted forms of $u(\cdot)$ and $y(\cdot)$.

Fig. 2.9 illustrates the transformation of input and output functions according to (2.4.1).

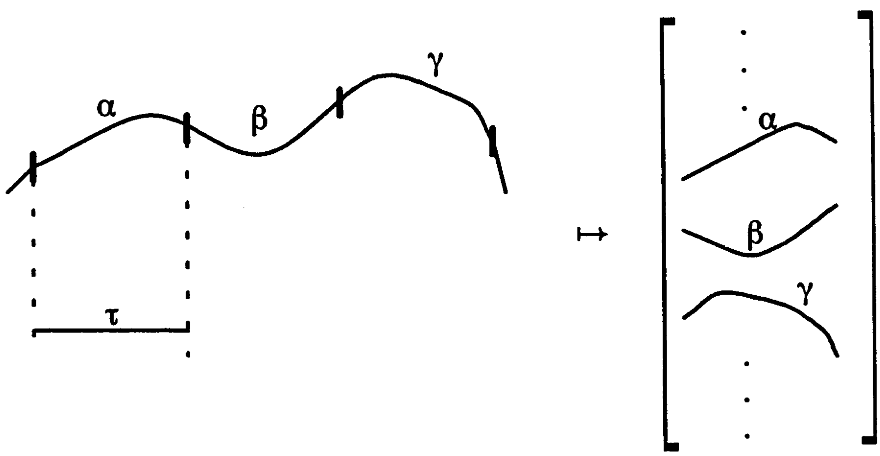


Figure 2.9. Function transformation in the Yamamoto's approach

Consider now a continuous-time system

$$\dot{x}(t)=Ax(t)+Bu(t)$$
$$y(t)=Cx(t) \qquad (2.4.2)$$

Applying (2.4.1) to the system (2.4.2) one can obtain the following time-invariant discrete-time system

$$x_{k+1} = e^{A\tau}\, x_k + \int_0^\tau e^{A(\tau-\varsigma)}\, B\, u_k(\varsigma)\, d\varsigma$$
$$y_k(\theta) = C\, e^{A\theta}\, x_k + \int_0^\theta C\, e^{A(\theta-\varsigma)}\, B\, u_k(\varsigma)\, d\varsigma, \qquad (2.4.3)$$

where $x_k = x(k\tau)$.
(Note that the intersample parameter θ enters as a parameter and not a time

variable).

The obtained system (2.4.3) has infinite–dimensional input/output spaces, but in return, is time–invariant. Thus, it can be easily connected with a digital controller without changing the time set.

Denoting state–space operators of the augmented control system (consisting of the lifted system (2.4.3) and a digital controller) as $\mathcal{A}$, $\mathfrak{B}$, $\mathcal{C}$ and $\mathfrak{D}$, Yamamoto associates with the original sampled–data system (consisting of the continuous–time system (2.4.2) and the controller connected to the system via a sampler and hold) the frequency response

$$\hat{\mathcal{T}}_Y(z) = \mathcal{C}(zI-\mathcal{A})^{-1}\,\mathfrak{B} + \mathfrak{D} \tag{2.4.4}$$

which relates z–transforms of the lifted input and output as follows:

$$\bar{y}(z) = \hat{\mathcal{T}}_Y(z)\,\bar{u}(z) \tag{2.4.5}$$

where $\bar{u}(z)$ and $\bar{y}(z)$ are z–transforms of input u and output y lifted according to (2.4.1):

$$\bar{u}(z) = \mathcal{Z}\{u_k(\theta)\}\bar{y}(z) = \mathcal{Z}\{y_k(\theta)\}$$

In terms of time–domain operators the inverse z–transform of $\hat{\mathcal{T}}_Y(z)$, denoted by $\mathcal{T}_Y$, can be written as

$$\mathcal{T}_Y = \mathrm{Sl}\ T\ \mathrm{Gl}, \tag{2.4.6}$$

where T is a sampled–data, no direct feedthrough system operator, Sl is an operator which lifts continuous–time signals slicing them into pieces of length τ, while Gl is an inverse operator which glues pieces of lifted continuous–time signal up. Clearly,

$$T = \mathrm{Gl}\ \mathcal{T}_Y\ \mathrm{Sl}. \tag{2.4.7}$$

2.4.2 Fast–Sampled/Lifted System and Yamamoto's System Relations

In the approach presented in this thesis we replace the sampled–data system by a system with fast sampled and lifted input, output and state. Fig. 2.10 illustrates the transformation of input, output and state functions when substituting the system associated with the operator $\mathcal{T}_{[N]}$ with N points for each vector element.

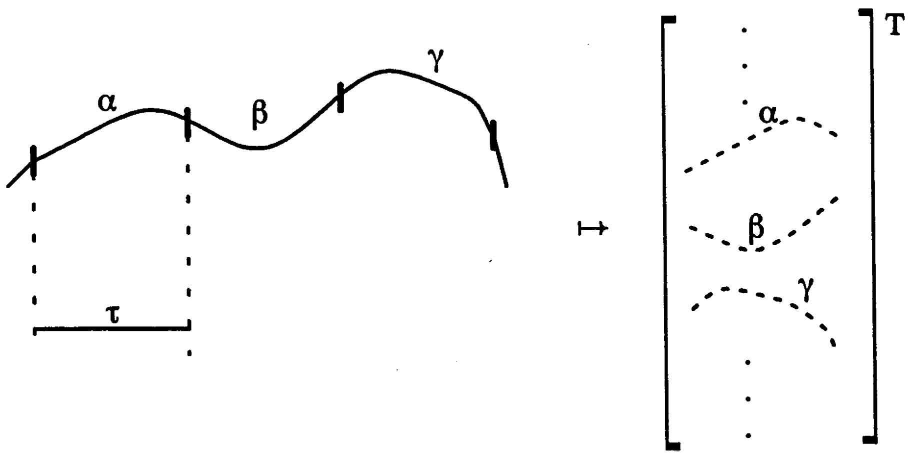


Figure 2.10. Function transformation in the approach of this thesis

Fig. 2.11 illustrates the transformations by operators T, Gl, Sl and $\mathcal{T}_Y$.

$$\begin{tikzcd}
l_2^{L_2[0,\tau)} \arrow[r, shift right, "Gl"'] \arrow[d, "\mathcal{T}_Y"'] & L_2[0,\infty) \arrow[l, shift right, "Sl"'] \arrow[d, "T"] \\
l_2^{L_2[0,\tau)} \arrow[r, shift right, "Gl"'] & L_2[0,\infty) \arrow[l, shift right, "Sl"']
\end{tikzcd}$$

Figure 2.11. Operators T, $\mathcal{T}_Y$, Gl and Sl

Note that the operators Sl and Gl are isomorphisms and T and $\mathcal{T}_Y$ map isomorphic spaces.

Zhang and Zhang (1994) studied properties of generalized lifted systems and showed that $\|\mathcal{T}_Y F\|_2 = \|T F\|_2$ where F is strictly causal and stable. Thus, it follows from Theorem 2.3 that the norm of $\tilde{\mathcal{T}}_{[N]} F$ is a good approximation of the norm of $\mathcal{T}_Y F$ for large N and any strictly causal and stable F. Namely, the following is true:

Theorem 2.4. Let $\mathcal{T}_{[N]}$ be the operator of the fast-sampled and lifted sampled-data system and let $\tilde{\mathcal{T}}_{[N]} F$ be constructed from $\mathcal{T}_{[N]}$ according to (2.3.3), where F is an operator satisfying the conditions of Corollary 2.1, and $\mathcal{T}_Y$ be the inverse z-transform of the operator $\hat{\mathcal{T}}_Y$ of the system obtained according to the Yamamoto's procedure. Then

$$\lim_{N \to \infty} \|\tilde{\mathcal{T}}_{[N]} F\|_2 = \|\mathcal{T}_Y F\|_2$$

We are however interested in comparing the behaviour at a particular frequency of the frequency responses of the lifted system (Yamamoto) and the fast sampled and lifted system. To do this, we need the following variant on Lemma 2.3.

Lemma 2.4. Let $H_{\tau/N}$, $S_{\tau/N}$ be as earlier defined and let T be the operator describing a linear, periodically time-varying system with impulse response $T_c(t,s)$, continuous in $t \geq s$, with $T_c(t+\tau,s+\tau)=T_c(t,s)$ for all $t \geq s$, and $\|T_c(t,s)\| \leq \alpha \exp[-\beta(t-s)]$ for some α, $\beta > 0$. (Here, $\|\cdot\|$ denotes the Euclidean norm). Consider periodic inputs to T with lifted description

$$u_Y = [\ \ldots,\ u,\ u\, e^{j\omega\tau},\ u\, e^{2j\omega\tau},\ \ldots\ ]$$

with $u(\theta)$ defined for $\theta=[0,\tau)$, $u(\theta) \in L_2[0,\tau)$; let

$$y_Y = [\ \ldots,\ y,\ y\, e^{j\omega\tau},\ y\, e^{2j\omega\tau},\ \ldots\ ]$$

denote the corresponding lifted output. Then with $\|u\|=1$,

$$\lim_{N \to \infty} \|H_{\tau/N} S_{\tau/N}\, y - y\|_2 = 0$$

with convergence uniform in $u(\cdot)$ in the sense that given $\varepsilon>0$, there exists $N_0(\varepsilon)$ such that for all $N \geq N_0$

$$\|H_{\tau/N} S_{\tau/N}\, y - y\|_2 < \varepsilon$$

independent of $u(\cdot)$.

Proof. For a unity norm $u(\cdot)$, we have

$$\begin{aligned}
&\lim_{N\to\infty} \|H_{\tau/N} S_{\tau/N}\, y - y\|_2 \\
&\leq \lim_{N\to\infty} \max_{\|u\|=1} \|H_{\tau/N} S_{\tau/N}\, y - y\|_2 \\
&= \lim_{N\to\infty} \|H_{\tau/N} S_{\tau/N}\, T - T\|_2 .
\end{aligned}$$

The last limit converges, according to Lemma 2.3, to zero. Thus, so does the limit

$\lim_{N\to\infty} \|H_{\tau/N} S_{\tau/N}\, y - y\|_2$ and this convergence is uniform in $u(\cdot)$. ❑

With this lemma in hand, we can now state the main result.

Theorem 2.5. Let T be the operator describing a periodic system with period τ. Suppose that T has no direct feedthrough term, i.e. T has a representation by an impulse response $T_c(t,s)$ which is continuous in $t \geq s$; suppose further that $\|T_c(t,s)\| \leq \alpha \exp[-\beta(t-s)]$ for some α, $\beta > 0$ and all $t \geq s$. Let $\hat{\mathfrak{T}}_Y(e^{j\omega\tau})$ denote the frequency response operator associated with the lifted version of the system defined via the Yamamoto approach, and let $\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})$ denote the frequency response of the fast-sampled and lifted version of T (with fast-sampling interval τ/N). Then for $\omega \in R_+$,

$$\lim_{N\to\infty} \|\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})\| = \|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\|.$$

Proof. Let ω be arbitrary but fixed. Let $u(\theta)$, $\theta \in [0,\tau)$, $u \in L_2[0,\tau)$ define an input to a Yamamoto-lifted description of T by

$$u_Y = [\, \ldots, u, u\, e^{j\omega\tau}, u\, e^{2j\omega\tau}, \ldots \,]$$

and let the corresponding output be

$$y_Y = [\, \ldots, y, y\, e^{j\omega\tau}, y\, e^{2j\omega\tau}, \ldots \,].$$

The norm $\|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\|$ at frequency ω is defined as

$$\|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\| = \sup_{u(\theta) \in L_2[0,\tau)} \|y\|_2 / \|u\|_2$$

Let ε be a small positive number. There exists a possibly complex $u_1 \in L_2[0,\tau)$ of unit norm such that

$$\|y_1\|_2/\|u_1\|_2 > \|\hat{\mathcal{T}}_Y(e^{j\omega\tau})\| - \varepsilon/3.$$

Of course, $y_1(\theta)$, $\theta \in [0,\tau)$ is the zeroth component of the output sequence associated with $u_1(\theta)$, $\theta \in [0,\tau)$.

Choose $N_0(\varepsilon)$ as follows. For all $N \geq N_0$ an approximation $u_2^{[N]}(\cdot)$ to $u_1(\cdot)$ depending on N, with $u_2^{[N]}(\cdot)$ constant on intervals of length τ/N and $u_2^{[N]}(\cdot)$ of unit $L_2[0,\tau)$ norm, ensures

$$\|y_2^{[N]}\|_2/\|u_2^{[N]}\|_2 > \|\hat{\mathcal{T}}_Y(e^{j\omega\tau})\| - 2\varepsilon/3.$$

Next, let $y_d^{[N]}$ denote the N-vector of equi-spaced values of $y_2^{[N]}$ samples being spaced by τ/N, and let $\bar{y}_2^{[N]}$ denote the associated piecewise constant signal. By replacing N_0 by a larger value if necessary that is nevertheless independent of $u_1(\cdot)$, we conclude from Lemma 2.4 (identifying y and $y_2^{[N]}$, and $H_{\tau/N}\, S_{\tau/N}\, y$ and $\bar{y}_2^{[N]}$) that

$$\|y_2^{[N]} - \bar{y}_2^{[N]}\|_2 < \varepsilon/3.$$

Then for all $N \geq N_0(\varepsilon)$

$$\|\bar{y}_2^{[N]}\|_2/\|u_2^{[N]}\|_2 > \|\hat{\mathcal{T}}_Y(e^{j\omega\tau})\| - \varepsilon.$$

Let $u_d^{[N]}$ be the N-vector of the N constant values assumed by $u_2^{[N]}(\cdot)$ over the sub-interval of $[0,\tau)$ of length τ/N. Consider the fast sampled, lifted system excited by the sequence

$$u_{fs} = [\ \ldots,\ u_d^{[N]},\ u_d^{[N]}\, e^{j\omega\tau},\ u_d^{[N]}\, e^{2j\omega\tau},\ \ldots\ ] \tag{2.4.8}$$

Then the response is precisely

$$y_{fs} = [\ \ldots,\ y_d^{[N]},\ y_d^{[N]}\, e^{j\omega\tau},\ y_d^{[N]}\, e^{2j\omega\tau},\ \ldots\ ] \tag{2.4.9}$$

Clearly also,

$$\|y_d^{[N]}\|/\|u_d^{[N]}\| = \|\bar{y}_2^{[N]}\|_2/\|u_2^{[N]}\|_2$$

(The norms on the left are ordinary Euclidean norms)

Hence for all $N \geq N_0(\varepsilon)$,

$$\|\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})\| = \sup_{u_d^{[N]}} \|y_d^{[N]}\|/\|u_d^{[N]}\| \geq \|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\| - \varepsilon$$

and so

$$\lim_{N\to\infty} \inf \|\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})\| \geq \|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\|. \tag{2.4.10}$$

We now aim to prove the reverse inequality. For any N, we can choose a unit norm $u_d^{[N]}$ so that the input sequence (2.4.8) to the fast-sampled lifted system results in the output sequence (2.4.9) such that

$$\|y_d^{[N]}\|/\|u_d^{[N]}\| = \|\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})\|.$$

Let $u_2^{[N]}(\theta)$, $\theta \in [0,\tau)$ be a time function defined over $[0,\tau)$ which is piecewise constant over intervals of length τ/N, with the constant values the entries of $u_d^{[N]}$. Let $\bar{y}_2^{[N]}(\theta)$, $\theta \in [0,\tau)$ be the corresponding time function associated with $y_d^{[N]}$. Further, let $y_2^{[N]}(\theta)$, $\theta \in [0,\tau)$ define the output sequence of the Yamamoto-lifted system with input sequence and output sequence

$$u_Y^{[N]} = [\ \ldots,\ u_2^{[N]},\ u_2^{[N]}\, e^{j\omega\tau},\ u_2^{[N]}\, e^{2j\omega\tau},\ \ldots\ ]$$

$$y_Y^{[N]} = [\ \ldots,\ y_2^{[N]},\ y_2^{[N]}\, e^{j\omega\tau},\ y_2^{[N]}\, e^{2j\omega\tau},\ \ldots\ ]$$

By Lemma 2.4 (noting the normalization of $\|u_2^{[N]}\|$ arising from $\|u_d^{[N]}\|=1$, it follows that given $\varepsilon>0$, there exists $N_0(\varepsilon)$, independent of $u_d^{[N]}$, such that for all $N \geq N_0(\varepsilon)$,

$$\|y_2^{[N]} - \bar{y}_2^{[N]}\|_2 < \varepsilon\, \|u_2^{[N]}\|.$$

Then for all $N \geq N_0(\varepsilon)$

$$\|\hat{\mathfrak{T}}_Y(e^{j\omega\tau})\| \geq \|y_2^{[N]}\|/\|u_2^{[N]}\| \geq \|\bar{y}_2^{[N]}\|/\|u_2^{[N]}\| - \varepsilon$$
$$= \|y_d^{[N]}\|/\|u_d^{[N]}\| - \varepsilon = \|\hat{\mathfrak{T}}_{[N]}(e^{j\omega\tau})\| - \varepsilon.$$

Hence

$$\|\hat{\mathcal{T}}_Y(e^{j\omega\tau})\| \geq \lim_{N\to\infty} \sup \|\hat{\mathcal{T}}_{[N]}(e^{j\omega\tau})\| \tag{2.4.11}$$

From (2.4.10) and (2.4.11) the desired equality follows. ❑

Remark 2.2. The problem considered in the theorem is nontrivial because while the norm is attained as the maximal singular value, the singular vectors (Schmidt pairs) are N–dependent, and one needs some sort of uniformity to prove the result.

Remark 2.3. The condition of no direct feedthrough term is essential and cannot be lifted. The desired uniformity of convergence cannot be guaranteed when there is a direct feedthrough term. The difference between the two situations is that in the case of systems without feedthrough terms, by the integration effect, the output is uniformly equicontinuous for $\|u\|_2 \leq 1$, and hence the distance between y and its sample–holded approximation is estimated uniformly for $\|u\|_2 \leq 1$.

2.5 Araki-Hagiwara-Ito Approach

Let us briefly review the Araki–Hagiwara–Ito approach. Let us define the signal set $\mathcal{X}_\phi$ as the set of all signals having finite power and consisting of sinusoidal components with equally spaced frequencies $\phi_m = \phi+2\pi m/\tau$. Assuming (without loss of generality) $-\pi/\tau<\phi\leq\pi/\tau$, the signal set $\mathcal{X}_\phi$ becomes as follows:

$$\mathcal{X}_\phi=\{x(t) \mid x(t)=\sum_{n=-\infty}^{\infty} x_n\, e^{j(\phi+2\pi n/\tau)t},\ \sum_{n=-\infty}^{\infty} \|x_n\|^2<\infty\}. \tag{2.5.1}$$

The importance of the signal set $\mathcal{X}_\phi$ lies in that a sampled–data system maps, in the steady state, a signal in $\mathcal{X}_\phi$ to a signal in the same set. From this fact, an operator $\hat{\mathcal{Q}}(j\phi)$ can be associated with the sampled–data system with the input signals restricted to within the signal set $\mathcal{X}_\phi$. According to Araki *et al.*

the operator is called the frequency response operator of the sampled-data system and its norm is called the frequency response.

To investigate the structure of the operator $\hat{\mathcal{Q}}$ let us consider an input x(t) from the signal set $\mathcal{X}_\phi$ and corresponding output y(t) which also belongs to $\mathcal{X}_\phi$ and is given by

$$y(t)=\sum_{n=-\infty}^{\infty} y_n\, e^{j(\phi+2\pi n/\tau)t}.$$

Then the operator matrix can be written in the form

$$\hat{\mathcal{Q}}(\phi) = \begin{bmatrix} & \vdots & \vdots & \vdots & \\ \cdots & Q_{-1,-1} & \cdots\ Q_{-1,0} & \cdots\ Q_{-1,1} & \cdots \\ & \vdots & \vdots & \vdots & \\ \cdots & Q_{0,-1} & \cdots\ Q_{0,0} & \cdots\ Q_{0,1} & \cdots \\ & \vdots & \vdots & \vdots & \\ \cdots & Q_{1,-1} & \cdots\ Q_{1,0} & \cdots\ Q_{1,1} & \cdots \\ & \vdots & \vdots & \vdots & \end{bmatrix}$$

where $Q_{m,n}$ is called the component-block of $\hat{\mathcal{Q}}(\phi)$.

Denoting by $\bar{x}$ and $\bar{y}$ the bi-directional series

$$\bar{x} = (\ldots x_{-2}^T, x_{-1}^T, x_0^T, x_1^T, x_2^T, \ldots)^T$$
$$\bar{y} = (\ldots y_{-2}^T, y_{-1}^T, y_0^T, y_1^T, y_2^T, \ldots)^T$$

the following holds:

$$\bar{y} = \hat{\mathcal{Q}}(\phi)\, \bar{x}. \tag{2.5.2}$$

Frequency response approximation is achieved by truncating the infinite-dimensional matrix $\hat{\mathcal{Q}}(\phi)$ and considering a finite-dimensional matrix $\hat{\mathcal{Q}}_{[2N+1]}(\phi)$ instead, consisting just of the central $(2N+1)\times(2N+1)$ component-blocks of $\hat{\mathcal{Q}}(\phi)$.

Thus, the output signal approximation can be written as

$$\tilde{y}_{[2N+1]}(t) = \sum_{n=-N}^{N} y_n\, e^{j(\phi+2\pi n/\tau)t} = \sum_{m=-N}^{N} \sum_{n=-N}^{N} Q_{n,m}\, e^{j(\phi+2\pi n/\tau)t}\, x_m .$$

The coefficients x_m are the input $x(t)$ Fourier series coefficients:

$$x_m = 1/\tau \int_{-\tau/2}^{\tau/2} x(t)\, e^{-j(\phi+2\pi m/\tau)t}\, dt$$

and the output approximation according to the Araki–Hagiwara–Ito approach is:

$$\tilde{y}_{[2N+1]}(t) = 1/\tau \sum_{n=-N}^{N} \sum_{m=-N}^{N} Q_{n,m} \int_{-\tau/2}^{\tau/2} x(\theta)\, e^{-j2\pi m\theta/\tau}\, d\theta\, e^{j(\phi+2\pi n/\tau)t} . \qquad (2.5.3)$$

As has been shown in Yamamoto and Araki (1994), the limit of the norms of the operators $\hat{\mathcal{Q}}_{[2N+1]}(\phi)$ converges to the norm of the Yamamoto operator described in Section 2.4 when $N\rightarrow\infty$. That means that, according to Theorem 2.4, the limits of the norms of the Araki–Hagiwara–Ito and fast-sampled/lifted systems are equivalent. Although the norms of both systems converge to the same function when $N\rightarrow\infty$, they are fundamentally different for finite values of N. In one of the approaches N is the number of different frequency sinusoids considered, while in the other approach N is the number of sample times over each nominal sampling period.

In the next section we will compare convergence rates and quality of approximation for fixed N for both approaches on examples.

2.6 Examples Study

2.6.1 Simple System Study

Let us consider a simple open-loop system shown on Fig. 2.12,

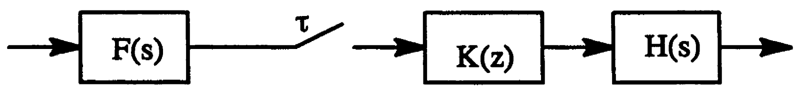


Figure 2.12. Simple sampled-data system

where H(s) is a zero-order hold, G(s) and K(z) are stable and F(s) is strictly proper. We are going to compare effectiveness of the approach given in Araki and Ito (1992, 1993, 1994), Araki *et al.* (1993, 1994), Hagiwara *et al.* (1993), Ito *et al.* (1993) and the fast-sampling/lifting approach described in this thesis.

2.6.1.1 Frequency Response Computation Using the Fast-Sampling and Lifting Approach

We can associate F(s) with its fast-sampled (τ/N) and lifted $N \times N$ version $\mathcal{F}_{[N]}(z)$.

Let us define the repeater:

$$E_1=(1 \;\; 1 \;\; 1 \; . \; . \; . \; 1)^T \quad \in \quad \mathbb{R}_{N\times 1} \tag{2.6.1a}$$

and decimator

$$E_2=(1 \;\; 0 \;\; 0 \; . \; . \; . \; 0) \in \quad \mathbb{R}_{1\times N}. \tag{2.6.1b}$$

Let $\mathcal{T}_{[N]}$ be the operator describing the fast-sampled with interval τ/N and lifted version of the system shown in Fig. 2.12. Then, denoting by $\hat{\mathcal{T}}_{[N]}$ the z-transform $\mathcal{Z}\{\mathcal{T}_{[N]}\}$,

$$\hat{\mathcal{T}}_{[N]}(z)=E_1K(z)\,E_2\;\;\mathcal{F}_{[N]}(z). \tag{2.6.2}$$

Bearing in mind the definition of singular values and the fact that the nonzero eigenvalues of the product of two matrices are invariant under change of the order of their multiplication, we can write for the limit of maximum singular values of the frequency response matrices $\hat{\mathcal{T}}_{[N]}(z)$ when $N \to \infty$:

$$\lim_{N\to\infty} \|\hat{\mathcal{T}}_{[N]}(z)\| = \lim_{N\to\infty} \bar{\lambda}^{1/2}\{\hat{\mathcal{T}}_{[N]}^{*}(z)\ \hat{\mathcal{T}}_{[N]}(z)\}$$
$$= \lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{F}_{[N]}^{*}(z)\ E_2^T\ K^{*}(z)\ E_1^T\ E_1\ K(z)\ E_2\ \mathcal{F}_{[N]}(z)\}$$
$$=\{K^{*}(z)\ K(z)\ \}^{1/2} \lim_{N\to\infty} \sqrt{N}\ \{\ [\mathcal{F}_{[N]}(z)\ \mathcal{F}_{[N]}^{*}(z)]_{1,1}\ \}^{1/2}. \tag{2.6.3}$$

Let us assume F(s) is a stable strictly proper one–state continuous–time element given in its state–space form: $F(s) = c(s-a)^{-1}b$.

Then, using (2.2.1–9) it is easy to obtain:

$$\{\ [\mathcal{F}_{[N]}(z)\ \mathcal{F}_{[N]}^{*}(z)]_{1,1}\ \}^{1/2}$$
$$= |bc/a|\ |\exp(a\tau/N) - 1|\ \sqrt{\exp(2a\tau) - 1}\ /\ \sqrt{(\exp(2a\tau/N) - 1)}$$
$$\div \sqrt{(z-\exp(a\tau))\quad (1/z - \exp(a\tau))}\ . \tag{2.6.4}$$

Calculating the limit of the N–dependent part of (2.6.3) we have:

$$\lim_{N\to\infty} \sqrt{N}\ (1 - \exp(a\tau/N))\ /\sqrt{1 - \exp(2a\tau/N)} = \sqrt{|a|\ \tau\ /\ 2}\ . \tag{2.6.5}$$

Now, we can re–write the frequency response magnitude of the system (2.6.3) as

$$\lim_{N\to\infty} \|\hat{\mathcal{T}}_{[N]}(z)\| = \{K^{*}(z)\ K(z)\ \}^{1/2}\ |bc|\ \sqrt{\tau}\ \sqrt{1 - \exp(2a\tau)}$$
$$\div \sqrt{2\ |a|\ (z - \exp(a\tau))\ (1/z - \exp(a\tau))}\ . \tag{2.6.6}$$

As we can see, the limit of the N–dependent part of (2.6.3) does not depend on z, and convergence to it occurs at the same rate at all frequencies. Also, observe (Fig. 2.13) that (2.6.3) converges to its limit (the frequency response of the system) (2.6.6) very fast.

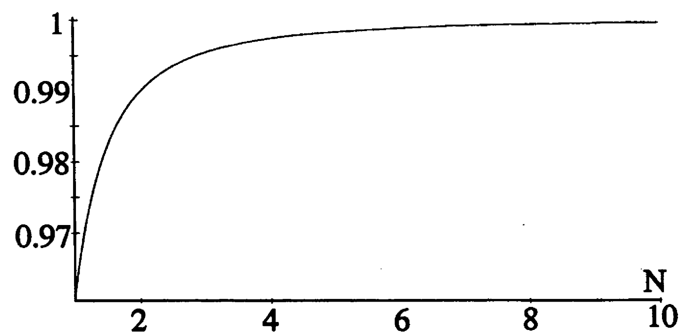


Figure 2.13. Rate of convergence of (2.6.3) to its limit (2.6.6), normalised to the unity limit value, with $a\tau = -1$.

2.6.1.2 Frequency Response Computation Using the Araki–Hagiwara–Ito Approach

In Araki and Ito (1992, 1993, 1994), Araki *et al.* (1993, 1994), Hagiwara *et al.* (1993), Ito *et al.* (1993), the frequency response of the sampled–data system is defined as the norm $\|\hat{\mathfrak{Q}}(j\phi)\|$ of the frequency response operator $\hat{\mathfrak{Q}}(j\phi)$, which matrix expression $\hat{\mathfrak{Q}}(j\phi)$ is, in turn, the limit of the matrices $\hat{\mathfrak{Q}}_{[2N+1]}(j\phi)$ when N goes towards infinity:

$$\hat{\mathfrak{Q}}(j\phi) = \lim_{N\to\infty} \hat{\mathfrak{Q}}_{[2N+1]}(j\phi) \tag{2.6.7}$$

and

$$\|\hat{\mathfrak{Q}}(j\phi)\| = \bar{\sigma}(\hat{\mathfrak{Q}}(j\phi)) = \lim_{N\to\infty} \bar{\sigma}(\hat{\mathfrak{Q}}_{[2N+1]}(j\phi)), \tag{2.6.8}$$

where the maximum singular value of $\hat{\mathfrak{Q}}_{[2N+1]}(j\phi)$ is easily calculated using the formula derived in Hagiwara *et al.* (1993):

$$\bar{\sigma}(\hat{\mathfrak{Q}}_{[2N+1]}(j\phi)) = \{K^*(e^{j\phi\tau})\, K(e^{j\phi\tau})\}^{1/2} |bc| \sqrt{1-\exp(-j\phi\tau)}$$

$$\times \{\sum_{r=-N}^{N} 1/(a^2+(\phi+2\pi r/\tau)^2)\}^{1/2} \{\sum_{r=-N}^{N} 1/(\phi+2\pi r/\tau)^2\}^{1/2}/\tau \tag{2.6.9}$$

2.6.1.3 Comparison of the Two Frequency Responses

As we can see, the N–dependent and frequency–dependent parts of (2.6.9) cannot be separated and the limits (2.6.7) and (2.6.8) converge to their limit values with different rates at different frequencies.

Also, taking the limit of (2.6.9), one can notice that the limits (2.6.8) and (2.6.3) converge to the same limit value (2.6.6), i.e. $\lim_{N\to\infty} \|\hat{\mathfrak{T}}_{[N]}(\exp(j\phi\tau))\| = \|\hat{\mathfrak{Q}}(j\phi)\|$. This verifies (for the simple system shown in Fig. 2.12) the general result given in Section 2.5, which states that although the two definitions of the sampled–data system frequency response are based on different considerations and approach the problem from different ways, both definitions lead to identical results.

Comparing the Araki–Hagiwara–Ito and Fast–sample and lift approaches, first observe that for any fixed frequency ϕ both sequences $\bar{\sigma}(\hat{\mathfrak{Q}}_{[2N+1]}(j\phi))$ and $\bar{\sigma}(\hat{\mathfrak{T}}_{[N]}(\exp(j\phi\tau)))$ are monotonically increasing. Then, for the ratio of the corresponding elements of the sequences we have

$$\bar{\sigma}(\hat{\mathfrak{T}}_{[2N+1]}(\exp(j\phi\tau)))/\bar{\sigma}(\hat{\mathfrak{Q}}_{[2N+1]}(j\phi))$$

$$=\tau|a|^{-1}[(e^{j\phi\tau}-e^{a\tau})(e^{-j\phi\tau}-e^{a\tau})]^{-1/2}[(1-e^{j\phi\tau})(1-e^{-j\phi\tau})]^{-1/2}(2N+1)^{1/2}$$

$$\times\ (1-e^{a\tau/(2N+1)})\ [(1-e^{2a\tau})/(1-e^{2a\tau/(2N+1)})]^{1/2}$$

$$\times\ \{\sum_{r=-N}^{N}[a^2+(\phi+r\omega)^2]^{-1}\}^{-1/2}\ \{\sum_{r=-N}^{N}[\phi+r\omega]^{-2}\}^{-1/2}\,. \qquad (2.6.10)$$

For the fixed frequency $\phi = \pi/\tau$ the ratio becomes

$$\bar{\sigma}(\hat{\mathfrak{T}}_{[2N+1]}(\exp(j\phi\tau)))/\bar{\sigma}(\hat{\mathfrak{Q}}_{[2N+1]}(j\phi))$$

$$=\pi\,(1-e^{2a\tau})^{1/2}\,(2N+1)^{1/2}\,[2\,|a|\,\tau\,(1+e^{a\tau})]^{-1}\,[(1-e^{a\tau/(2N+1)})/(1+e^{a\tau/(2N+1)})]^{1/2}$$

$$\times\{\sum_{r=-N}^{N}[a^2\tau^2+\pi^2\,(2r+1)^2]^{-1}\}^{-1/2}\ \{\sum_{r=-N}^{N}[2r+1]^{-2}\}^{-1/2}\,. \qquad (2.6.11)$$

Computer simulations show that the ratio (2.6.11) is greater than one for any choice of $a<0$ (i.e. for any choice of $F(s)$ and $K(z)$) and for any $N \geq 0$. That means that $\bar{\sigma}\,(\hat{\mathcal{T}}_{[2N+1]}(\exp(j\phi\tau)))$ converges to its limit faster than $\bar{\sigma}(\hat{\mathcal{Q}}_{[2N+1]}(j\phi))$. Also, the fact that $\bar{\sigma}\,(\hat{\mathcal{T}}_{[N]}(\exp(j\phi\tau)))$ converges with the same rate irrespective of the frequency is another favourable feature of the Fast-sample and lift approach applied to the system of Fig. 2.12, which might be very useful in many applications.

2.6.2 *Combined System Study*

Now, let us consider a more complicated sampled-data system shown on Fig. 2.14, where $H(s)$ is a zero-order hold, $F(s)$, K, $G(s)$ are stable and $G(s)$ and $F(s)$ are strictly proper. Also, in order to simplify calculations, let us assume that $F(s)$ is first order.

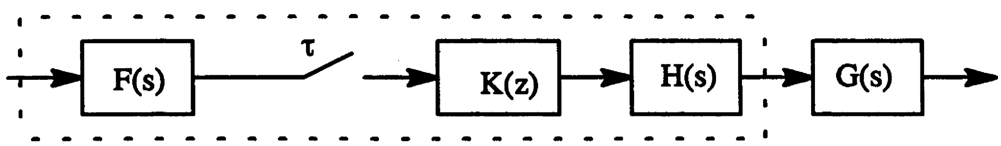


Figure 2.14. More complicated sampled-data system

The frequency response magnitude of the system is:

$$\lim_{N\to\infty} \|\hat{\mathcal{T}}_{[N]}(e^{j\omega\tau})\| = \lim_{N\to\infty} \bar{\lambda}^{1/2}\{\hat{\mathcal{T}}_{[N]}^*(e^{j\omega\tau})\,\hat{\mathcal{T}}_{[N]}(e^{j\omega\tau})\}$$

$$= \lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{F}_{[N]}^*(e^{j\omega\tau})E_2^T\, K^*(e^{j\omega\tau})E_1^T\mathcal{G}_{[N]}^*(e^{j\omega\tau}) \times \mathcal{G}_{[N]}(e^{j\omega\tau})E_1\, K(e^{j\omega\tau})E_2\mathcal{F}_{[N]}(e^{j\omega\tau})\}$$

$$=\{K^*(e^{j\omega\tau})\,K(e^{j\omega\tau})\,\}^{1/2} \lim_{N\to\infty}\{\sum_{r,s=1}^{N}[\mathcal{G}_{[N]}^*(e^{j\omega\tau})\,\mathcal{G}_{[N]}(e^{j\omega\tau})]_{r,s}\}^{1/2} \times\{\,[\mathcal{F}_{[N]}(e^{j\omega\tau})\,\mathcal{F}_{[N]}^*(e^{j\omega\tau})]_{1,1}\,\}^{1/2}. \qquad (2.6.12)$$

From (2.6.4) and (2.6.5) we can conclude that

$$\{\,[\mathcal{F}_{[N]}(e^{j\omega\tau})\,\mathcal{F}_{[N]}^*(e^{j\omega\tau})]_{1,1}\,\}^{1/2} = \underline{O}(N^{-1/2}). \qquad (2.6.13)$$

From Theorem 2.2 and Remark 2.1 it follows that the limit

$\lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{G}_{[N]}^*(e^{j\omega\tau})\,\mathcal{G}_{[N]}(e^{j\omega\tau})\}$ does exist and equals $|G(j\omega)|$ for small enough sampling period τ.

Symbolic computer experiments show that

$$\lim_{N\to\infty} \{ \sum_{r,s=1}^{N} [\mathcal{G}_{[N]}^*(e^{j\omega\tau})\,\mathcal{G}_{[N]}(e^{j\omega\tau})]_{r,s} /N\}^{1/2} = \lim_{N\to\infty} \bar{\lambda}^{1/2}\{\mathcal{G}_{[N]}^*(e^{j\omega\tau})\,\mathcal{G}_{[N]}(e^{j\omega\tau})\}. \tag{2.6.14}$$

Now, the frequency response of the system can be re-written as

$$\lim_{N\to\infty} \|\hat{\mathcal{F}}_{[N]}(e^{j\omega\tau})\|$$

$$= \{ K^*(e^{j\omega\tau})\, K(e^{j\omega\tau}) \}^{1/2} \lim_{N\to\infty} \sqrt{N}\{ [\mathcal{F}_{[N]}(e^{j\omega\tau})\, \mathcal{F}_{[N]}^*(e^{j\omega\tau})]_{1,1} \}^{1/2} \; |G(j\omega)| \; . \tag{2.6.15}$$

Comparing (2.6.15) and (2.6.3) we can conclude that the frequency response magnitude of the system shown on Fig. 2.14 can be calculated as a product of the frequency response of the simple subsystem (in dashed box or as in Fig. 2.12) and a modulus of the remaining continuous-time part of the system. This extremely powerful feature allows one to simplify calculations dramatically since to calculate the frequency response of the complex system it is enough to calculate the frequency response of the simple sampled-data system of Fig. 2.12.

2.6.3 Closed-Loop System Example

We now present a practical example to confirm the applicability (especially ease of use) of the approach to the closed-loop sampled-data control system. This example was first described by Salehi (1985) and then studied by McFarlane and Glover (1990).

The system considered in this example is a satellite with two highly flexible solar arrays attached. The model represents the transfer function from the torque applied to the roll axis of the satellite to the corresponding satellite roll angle. In order to keep the model simple, only a rigid body mode and a single flexible mode are included, resulting in a four state model.

The state-space matrices of the plant are given by

$$A_p = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega^2 & -2\zeta\omega \end{pmatrix}$$

$$B_p^T = [0 \quad 1.7319\times10^{-5} \quad 0 \quad 3.7859\times10^{-4}]$$

$$C_p = [1 \quad 0 \quad 1 \quad 0]$$

$$D_p = 0$$

where $\omega = 1.539$ rad/sec is the frequency of the flexure mode, and $\zeta = 0.003$ is the flexural damping ratio. The open-loop poles of the plant are at $-0.0046\pm1.5390j$, 0, 0.

Design of the controller to satisfy certain objectives given in Salehi (1985) and McFarlane and Glover (1990) results in the continuous controller with the following state-space representation (McFarlane and Glover, 1990):

$$A_c = \begin{pmatrix} -23.7320 & 1.0000 & -23.7320 & 0 \\ -4.4097 & -0.2308 & -4.3847 & -0.0212 \\ -8.4894 & 0 & -8.4894 & 1.0000 \\ -19.3781 & -5.0461 & -21.1999 & -0.4724 \end{pmatrix}$$

$$B_c^T = 10^4 \times [4.7464 \quad 0.8750 \quad 1.6979 \quad 3.7242]$$

$$C_c = [1 \quad 6.6643 \quad 0.2780 \quad 0.6117]$$

$$D_c = 0.$$

The controller is open-loop stable.

A discrete-time controller with sampling time τ=0.1858 sec. is obtained, by finding the zero-order hold equivalent of the open-loop controller. (This procedure, although it appears satisfactory in this case, can be subject to criticism on the grounds that it does not seek directly to have the continuous-time closed-loop closely approximated by the sampled-data closed loop as in Keller and Anderson, 1991, 1992). The discrete-time controller can be described by the following state-space form:

$$A=\begin{pmatrix} 0.3016 & 0.1162 & -0.6821 & -0.1067 \\ -0.1211 & 0.9479 & -0.1130 & -0.0223 \\ -0.3146 & -0.1106 & 0.6611 & 0.1325 \\ -0.3358 & -0.9095 & -0.6477 & 0.8197 \end{pmatrix}$$

$B^T = 10^3 \times [1.4094 \quad 0.2325 \quad 0.6098 \quad 0.4203]$, $C = C_c$, $D = 0$.

The antialiasing filter introduced into the loop has transfer function $F(s) = 5.5/(s+5.5)$. The elements of the system are interconnected as in Fig. 2.15.

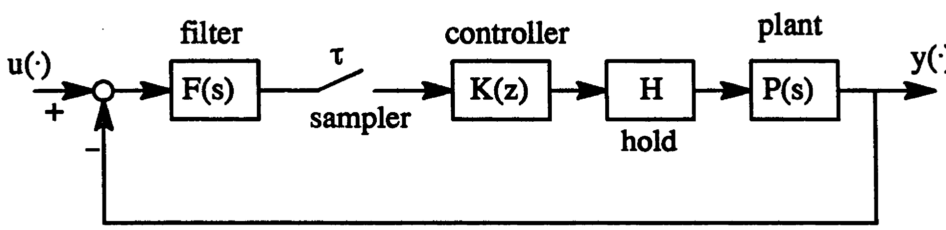


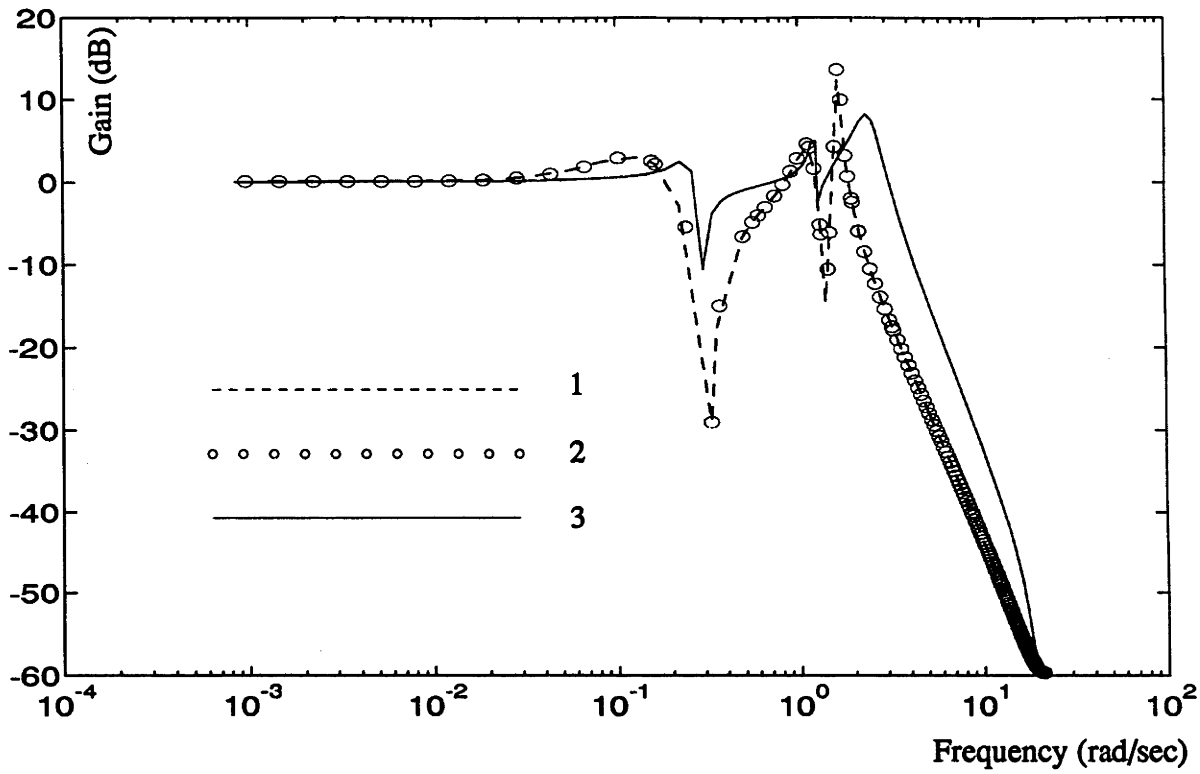
Figure 2.15. Sampled–data control system

Fig. 2.16 shows frequency response magnitudes of the closed-loop system. It is clearly seen that $\|\hat{\mathfrak{T}}_{[3]}(e^{j\phi\tau})\|$ approximates the frequency response magnitude of the hybrid closed-loop system ($\|\hat{\mathcal{Q}}(j\phi)\| = \lim_{N\to\infty} \|\hat{\mathfrak{T}}_{[N]}(e^{j\phi\tau})\|$) by far better then $\|\hat{\mathcal{Q}}_{[3]}(j\phi)\|$. (Here, $\|\hat{\mathfrak{T}}_{[3]}(e^{j\phi\tau})\|$ and $\lim_{N\to\infty} \|\hat{\mathfrak{T}}_{[N]}(e^{j\phi\tau})\|$ were calculated as was described in Section 2.3 and $\|\hat{\mathcal{Q}}_{[3]}(j\phi)\|$ and $\|\hat{\mathcal{Q}}(j\phi)\|$ were computed according to the method due to Araki, Hagiwara and Ito.) This shows that the approach suggested in this thesis may on occasion give better results than the approach developed in Araki and Ito (1992, 1993, 1994), Araki *et al.* (1993, 1994), Hagiwara *et al.* (1993) and Ito *et al.* (1993).

2.6.4 Output Approximation Comparison

Consider a sampled–data system defined by a linear periodically time–varying operator with associated causal impulse response $h(t,s) = 10^{-16}\,(t-s)^{20} \exp(s-t)$.

Fig. 2.17 depicts the output signals of the system with $u(t) = \sin t$ and $\tau = 1$. Comparing the output signal and its approximations obtained according to the fast-sampling/lifting approach and the Araki–Hagiwara–Ito approach



1 frequency response magnitude of the hybrid closed loop $(\lim_{N\to\infty}\|\hat{\mathcal{T}}_{[N]}\|$ or $\|\hat{\mathcal{Q}}\|)$

2 approximation of the frequency response magnitude of the hybrid closed loop obtained by fast sampling and lifting $(\|\hat{\mathcal{T}}_{[3]}(e^{j\phi\tau})\|)$

3 approximation of the frequency response magnitude of the hybrid closed loop obtained using the approach developed by Araki, Hagiwara and Ito $(\|\hat{\mathcal{Q}}_{[3]}(j\phi)\|)$

Figure 2.16. Frequency response of the hybrid closed loop and its approximations

with the same N for the fast-sampling coefficient and the number of harmonics, one sees again that the former method gives better approximation for finite N (even though both results converge to the actual output with N approaching infinity). In our example for N=3 the fast-sampling/lifting approach gives good approximation of the actual output. The approximation obtained using the Araki-Hagiwara-Ito approach with N=3 is a sum of 3 sinusoids with the amplitudes so small that the sum is indistinguishable from zero on the same graph and in no degree resembles the actual output it is supposed to approximate.

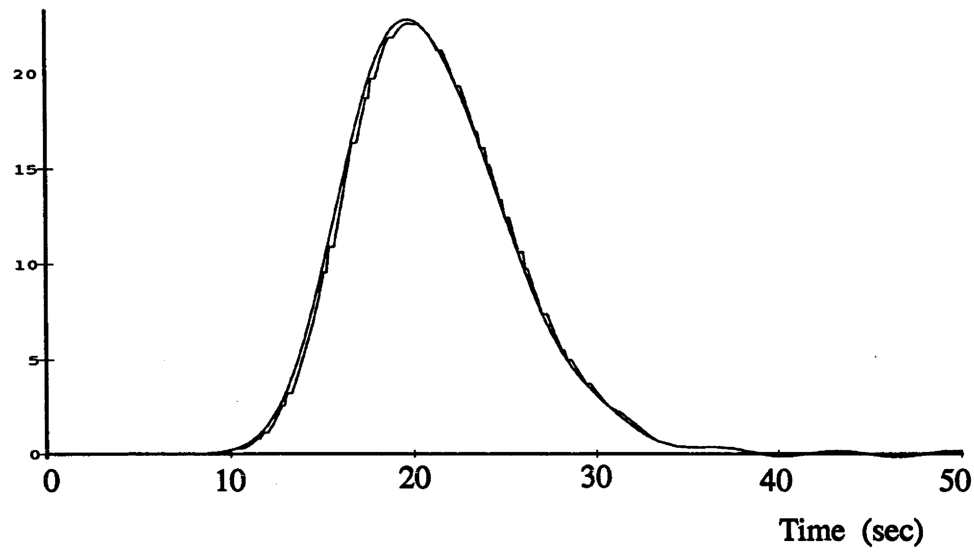


Figure 2.17. Output signal of the hybrid system and its fast-sampled/lifted approximation with N=3

2.7 Conclusions

A new approach to establish a frequency-domain paradigm for the sampled-data control systems has been presented. The key idea is to associate a sampled-data system with a discrete system obtained from the original sampled-data system by very fast sampling followed by lifting to convert the sampled multi-rate system to a single-rate one.

One of the most important questions related to the problem is how to compute this frequency response. It is easy to compute approximately. Similar question arises in the two other sampled-data system frequency response theories. (Araki and Ito, 1992, 1993, 1994; Araki *et al.*, 1993, 1994; Hagiwara *et al.*, 1993; Ito *et al.*, 1993; Yamamoto, 1993; Yamamoto and Araki, 1994; Yamamoto and Khargonekar, 1993). All calculation procedures rely on approximation which in turn is based on the truncation of infinite-dimensional operators at some finite dimension and/or γ-iteration. (Yamamoto and Khargonekar, 1993; Yamamoto and Araki, 1994)

The examples studied in this Chapter compared computational procedures of the approaches and show the clear benefit of fast-sampling and lifting. Examples suggest that good finite approximation requires a large number of sinusoidal frequencies in the approach described in Araki and Ito (1992, 1993, 1994), Araki *et al.* (1993, 1994), Hagiwara *et al.* (1993), Ito *et al.* (1993), while the integer N chosen in the fast sampling approach can assume just modest values.

Also, it was shown that for a simple open-loop sampled-data system, all three approaches converge to similar frequency-response formulae, although the approach based on fast-sampling/lifting may converge faster and uniformly.

God made the integers,
all the rest is the work of man.

Leopold Kronecker
Jahrsberichte der Deutschen
Mathematiker Vereinigung bk. 2

Chapter 3

Optimum Realizations of Sampled-Data Controllers for FWL Sensitivity Minimization

3.1 Introduction

It is well known that a desired controller's transfer function can be implemented by any one of an infinite set of realizations of the controller. Though all these realizations are in principle equivalent since they yield the same transfer function, they have different numerical properties due to finite wordlength effects when they are implemented by a digital device. A problem of great importance is to find the realization of the controller which achieves the best performance of the closed-loop system, i.e. gives the best approximation of the ideal closed loop behaviour.

An outline of this Chapter is as follows. In Section 3.2 we establish the definitions of sensitivity "functions" (operators) and L_2 sensitivity measure of a closed loop system. In Section 3.3 we study the finite-wordlength-optimal realization minimizing a measure of the sensitivity of the closed-loop operation with respect to controller coefficient errors. (No claim is made about FWL roundoff noise effects). The existence and uniqueness of an optimal solution are established. A recursive algorithm for obtaining the optimal solution is given. In Section 3.4 we implement the recursive algorithm. Two numerical examples to confirm theoretical results are given in Section 3.5 followed by some concluding remarks in Section 3.6.

3.2 Sensitivity Measure of a Realization

First consider a discrete linear time-invariant multi input, multi output controller having a transfer function K(z), which can be expressed in terms of matrices A, B, C, D of a minimal state-space realization as follows:

$$K(z)=C(zI_R - A)^{-1}B + D, \tag{3.2.1}$$

where $A\in\mathbb{R}^{R\times R}$, $B\in\mathbb{R}^{R\times L}$, $C\in\mathbb{R}^{M\times R}$, $D\in\mathbb{R}^{M\times L}$, $K\in\mathbb{C}^{M\times L}$.

Clearly, if matrices A, B, C, D satisfy (3.2.1), then for any similarity transformation T matrices $T^{-1}AT$, $T^{-1}B$, CT, D satisfy (3.2.1) as well. This means that there exist an infinite number of representations of the system. All these representations are equivalent insofar as they yield the same transfer function. However, different realizations have different numerical properties such as sensitivity to coefficient errors and propagation of signal roundoff errors. This means that in the finite precision case all these realizations are no longer equivalent. In practice it is impossible to realize matrices A, B, C, D exactly due to finite word length (FWL) constraints. As a result, the transfer function given by (3.2.1) and the transfer function with matrices A, B, C, D replaced by their FWL versions are different. Since different FWL realizations have different sensitivities, our task is to search for those realizations that minimize the sensitivity in some appropriate measure reflecting the overall control objective.

In order to define such a measure, we shall use the derivatives of elements of the controller transfer function matrix at an arbitrary but fixed value of z with respect to the elements of the matrices A, B, C of the realization:

$$\partial k_{m,l}/\partial a_{r,q}=g_{m,r}\, f_{q,l},\quad \partial k_{m,l}/\partial b_{r,i}=g_{m,r}\, \delta_{l,i},\quad \partial k_{m,l}/\partial c_{j,r}=f_{r,l}\, \delta_{m,j}, \tag{3.2.2}$$

where a, b, c, k, g, f are elements of the matrices A, B, C, K, G, F respectively, with

$$G=C(zI_R - A)^{-1} \in \mathbb{C}^{M\times R}, \tag{3.2.3a}$$

$$F=(zI_R - A)^{-1}B \in \mathbb{C}^{R\times L}, \tag{3.2.3b}$$

$l,i=1,2,\ldots,L$, $m,j=1,2,\ldots,M$, $r,q=1,2,\ldots,R$ and δ is the Kronecker delta. Note that the matrix D is coordinate independent and has nothing to do with the optimal realization problem.

Our major goal is to find the optimal implementation of the controller for achieving the best performance of the closed-loop system where the controller is implemented with FWL. "Best performance" can mean many things. As made more precise below, we shall consider the accuracy of implementing the input-output operator for the closed loop.

Consider a hybrid closed loop where the plant is continuous-time and the controller is discrete-time. (Such a configuration represents the usual situation). This closed loop is drawn in Fig. 3.1, where Π stands for the $L \times M$ continuous-time plant, K for the $M \times L$ discrete controller, Φ for the strictly proper stable antialiasing filter, Σ for the sampler with the sampling period τ and H for the hold element, here assumed to be a zero-order hold. (In the multivariable situation, Φ, Σ and H are diagonal operators).

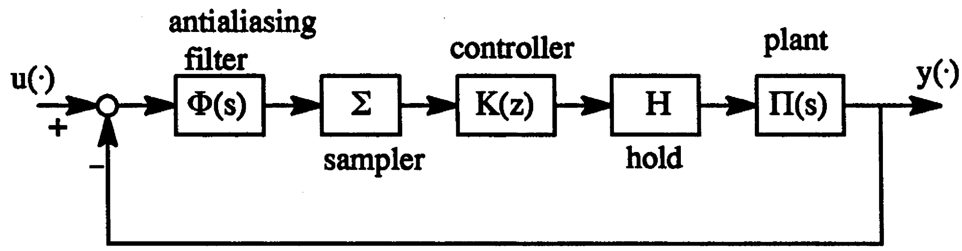


Figure 3.1: The closed-loop system

First of all we need to define a sensitivity measure of the closed-loop operator with respect to errors in the realization A, B, C of the controller. Then the problem of minimizing of this measure will arise.

Earlier works (Gevers and Li, 1993, Li and Gevers, 1990a, 1990b, 1991, 1993, Liu and Skelton, 1990, Liu *et al.*, 1992, Williamson and Kadiman, 1989) looked at a purely discrete-time problem and it was possible (and easier) to deal just with frequency domain quantities. However, in order to take into consideration intersample behaviour of the hybrid closed loop we need to work in the time domain.

Closed-loop operator and sensitivity with respect to a controller parameter

We shall assume that the sampling interval is such that no unstabilizable or undetectable modes are introduced by the sampling operator and that the closed loop is stable. Unstabilizable or undetectable modes can only occur for nongeneric $\Pi(s)$, and even then, only for isolated choices of sampling interval (Francis and Georgiou, 1988). Stability of the closed-loop system means that with zero input, any nonzero initial state decays to zero exponentially fast, and $u(\bullet) \in L_p^L[0,\infty)$ implies $y(\bullet) \in L_p^L[0,\infty)$ for all $p \in [1,\infty]$, see (Francis and Georgiou, 1988). The closed loop is defined by a linear periodically time-varying operator $\mathcal{H}$ with associated causal impulse response $\mathcal{H}(t,s)$, such that

$$\bar{y}(t) = \int_{-\infty}^{t} \mathcal{H}(t,s)\, \bar{u}(s)\, ds \tag{3.2.4}$$

$$\mathcal{H}(t+\tau,s+\tau) = \mathcal{H}(t,s) \tag{3.2.5}$$

The stability condition is expressed by

$$\|\mathcal{H}(t,s)\|_F \leq \alpha \exp\,[-\beta(t-s)] \tag{3.2.6}$$

for some $\alpha > 0$, $\beta > 0$, with the subscript F denoting the Frobenius norm so that

$$\|A\|_F = [\mathrm{tr}\,(A^T A)]^{1/2}.$$

A composition of two stable operators is stable.
Formally, with minimal abuse of notation, we can write

$$\mathcal{H} = \Pi H K \Sigma \Phi\, (I + \Pi H K \Sigma \Phi)^{-1}, \tag{3.2.7}$$

where Π, H, K, Σ, Φ are the operators corresponding to the blocks shown in Fig. 3.1, and then the derivative of $\mathcal{H}$ with respect to an element in a realization of K can be formally written as

$$\partial \mathcal{H} / \partial \alpha \quad = \mathcal{V}\ \partial K/\partial \alpha \quad \mathcal{W} \tag{3.2.8a}$$

where

$$\mathcal{V} = (I_L + \Pi H K \Sigma \Phi)^{-1}\ \Pi H\ , \tag{3.2.8b}$$

$$\mathcal{W} = \Sigma\Phi(I_L + \Pi H K \Sigma \Phi)^{-1}. \tag{3.2.8c}$$

Because of the stability of the closed-loop, $\mathcal{V}$ and $\mathcal{W}$ map $l_p^M(Z_+)$ into $L_p^L[0,\infty)$ and $L_p^L[0,\infty)$ into $l_p^L(Z_+)$ for all $p \in [1,\infty]$, including of course $p = 2$, respectively. Moreover, the mappings are causal.

The derivative (3.2.8a) of the closed-loop operator $\mathcal{H}$ can be represented as in Fig. 3.2.

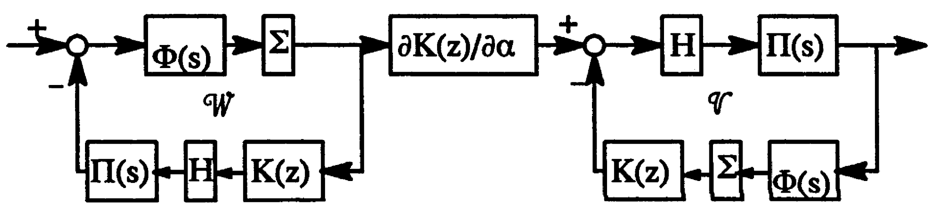


Figure 3.2: Representation of derivative of closed-loop operator with respect to parameter α of controller

The representation of Fig. 3.2 has a deficiency that should be remedied. If $K(z)$ is open-loop unstable, $\partial K(z)/\partial \alpha$ has this property also, see (3.2.2) and (3.2.3); yet if the closed loop is stable, one would expect that the operator $\partial\mathcal{H}/\partial\alpha$ should also have this property. This is in fact so. The representation of Fig. 3.2 can be replaced by one (with lower overall state dimension) which is stable. This is done as follows:

Recognize that if $\alpha = a_{i,j}$, then (with some minor abuse of notation)

$$\partial\mathcal{H}/\partial a_{i,j}$$

$$= (I_L + \Pi H K \Sigma \Phi)^{-1}\ \Pi H C(zI_R - A)^{-1}\ e_i\ e_j^T\ (zI_R - A)^{-1} B \Sigma \Phi (I_L + \Pi H K \Sigma \Phi)^{-1}$$

$$= \mathcal{V}_A\ e_i\ e_j^T \mathcal{W}_A \tag{3.2.9a}$$

where $\mathcal{V}_A$ and $\mathcal{W}_A$ (depicted in Fig. 3.3) are stable operators differing marginally from $\mathcal{V}$ and $\mathcal{W}$ in terms of the points where the loop input is introduced or the loop output is taken from. Also, the range of $\mathcal{W}_A$ and the domain of $\mathcal{V}_A$ are discrete-time signals: $\mathcal{W}_A$ and $\mathcal{V}_A$ are bounded operators from $L_p^L[0,\infty)$ to $l_p^R(Z_+)$ and $l_p^R(Z_+)$ to $L_p^L[0,\infty)$ $\forall p \in [1,\infty]$. Similarly,

$$\partial\mathcal{K}/\partial b_{i,j} = \mathcal{V}_A\, e_i\, e_j^T\, \mathcal{W} \tag{3.2.9b}$$

$$\partial\mathcal{K}/\partial c_{i,j} = \mathcal{V}\, e_i\, e_j^T\, \mathcal{W}_A \tag{3.2.9c}$$

Because $\mathcal{V}$, $\mathcal{W}$, $\mathcal{V}_A$ and $\mathcal{W}_A$ are all stable operators, the operators on the right in (3.2.9) are all stable.

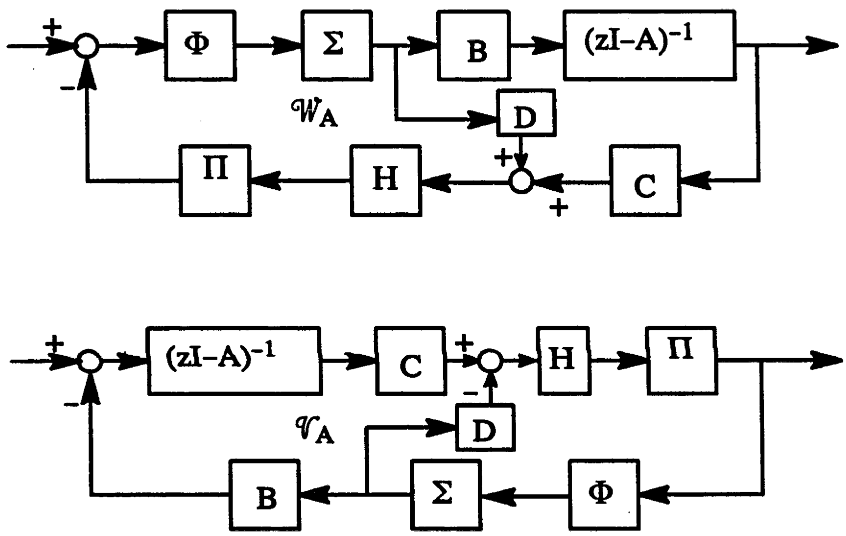


Figure 3.3: Modification of operators $\mathcal{V}$ and $\mathcal{W}$ used in alternative construction of the derivatives of the closed-loop operator

In formulating a sensitivity "function" (here, more properly, operator) associated with a realization of a system, it is conventional to organize the matrix calculations slightly differently; one picks a particular entry of $\mathcal{K}$, $\mathcal{K}_{k,l}$ say, and constructs the matrices $\partial\mathcal{K}_{k,l}/\partial A$, $\partial\mathcal{K}_{k,l}/\partial B$, $\partial\mathcal{K}_{k,l}/\partial C$, where the i,j entry of $\partial\mathcal{K}_{k,l}/\partial A$ is $\partial\mathcal{K}_{k,l}/\partial a_{i,j}$. In the light of (3.2.9) it is clear that

$$\partial\mathcal{K}_{k,l}/\partial A = \mathcal{V}_A^T\, e_k\, e_l^T\, \mathcal{W}_A^T \tag{3.2.10a}$$

$$\partial\mathcal{K}_{k,l}/\partial B = \mathcal{V}_A^T\, e_k\, e_l^T\, \mathcal{W}^T \tag{3.2.10b}$$

$$\partial \mathcal{K}_{k,l} / \partial C = \mathcal{V}^T e_k e_l^T \mathcal{W}_A^T \tag{3.2.10c}$$

Here, $\mathcal{V}^T$ is not the adjoint operator of $\mathcal{V}$; rather, $\mathcal{V}^T$ is defined by the condition $e_i^T \mathcal{V}^T e_j = e_j^T \mathcal{V} e_i =$ (j–i) component of $\mathcal{V}$; thus $\mathcal{V}^T$ is $\mathcal{V}$ with elements reorganized.

Let us sum up results to this point:

Theorem 3.1

Consider the closed–loop system depicted in Fig. 3.1, comprising strictly proper (nonzero) stable antialiasing filter $\Phi(s)$, sampler Σ with sampling interval τ, (nonzero) discrete–time controller K(z), zero–order hold H and (nonzero) plant $\Pi(s)$. Suppose that τ is such that no unstabilizable or undetectable modes are introduced by the sampling operator, and the closed loop, defined by a periodically time–varying impulse response $\mathcal{K}(t,s)$, is stable. Let the controller have a minimal state–variable realization $C(zI - A)^{-1}B + D$, and let stable closed–loop operators $\mathcal{V}$, $\mathcal{W}$, $\mathcal{V}_A$ and $\mathcal{W}_A$ be defined as depicted in Figs. 3.2 and 3.3. Then the sensitivity functions of $\mathcal{K}$ with respect to the elements of A, B, C in the controller realization are given by (3.2.10) for k,l = 1, 2, . . . , L.

A numerical sensitivity measure

In order to determine a single numerical measure of sensitivity, we will use a norm associated with the sensitivity "functions". This is not an induced norm, but rather a norm associated with the impulse response representation of a stable operator, viewing it simply as a function of time in two variables. We confine attention to a matrix impulse response $\mathcal{U}(t,s)$ defined in the half plane $s \leq t$, with the periodicity property $\mathcal{U}(t+\tau, s+\tau) = \mathcal{U}(t,s)$ and with the (exponential) stability property $\|\mathcal{U}(t,s)\|_F \leq \alpha \exp[-\beta(t-s)]$ for some positive α, β. (We remark that in Francis and Georgiou (1988), there is also a departure from the case of induced norms in defining norms for a stable, periodically time–varying linear system.) The norm is

$$\|\mathcal{U}\|_2 = \{\int_0^\tau dt \int_{-\infty}^t \|\mathcal{U}(t,s)\|_F^2 ds\}^{1/2}.$$

Notice that in view of the periodicity property, the norm reflects all values assumed by $\mathcal{U}(t,s)$ in $-\infty < s \leq t < \infty$, even though the integration with respect to t only extends over $[0,\tau]$. Note also the alternative expression obtained by changing the order of integration:

$$\|\mathcal{U}\|_2 = \{\int_0^\tau ds \int_s^\infty \|\mathcal{U}(t,s)\|_F^2 dt\}^{1/2}.$$

Now we can define the sensitivity measure of the closed–loop operator with respect to the realization A, B, C of the controller:

Definition 3.1.
The sensitivity measure M_2 of the closed–loop operator with respect to the realization of the controller is the sum of the squares of the L_2 norms of sensitivity operators of the closed loop with respect to the matrices A, B, and C of the realization of the controller:

$$M_2 = \Sigma_{k,l}\{||\partial\mathcal{G}_{k,l}/\partial A||_2^2 + ||\partial\mathcal{G}_{k,l}/\partial B||_2^2 + ||\partial\mathcal{G}_{k,l}/\partial C||_2^2\}. \tag{3.2.13}$$

It is worth noting how the measure differs from those employed in earlier optimum realization problems. First, the measure is intrinsically a time–domain one, rather than a frequency domain one. Second, for reasons of mainly analytic convenience, in most earlier optimum realization work frequency domain L_2 norms of sensitivity functions related to B and C are used, while a frequency domain L_1 norm is used for the sensitivity function related to A. The frequency domain L_2 norms have some parallel (through Parseval's theorem) with our time–domain L_2 norms. Frequency domain L_1 norms of course are virtually unrelatable to a time–domain norm. References Gevers and Li (1993) and Li and Gevers (1991) use a frequency domain L_2 norm for the sensitivity function related to A, and are closest in spirit to this work, even though these references use a discrete–time model for the system. Reference Perkins *et al.* (1990) also uses an L_2 norm in an optimal filter realization problem.

3.3 Optimal FWL Realizations

The numerical measures of sensitivity defined above depend on the particular realization of the controller. We shall now clarify the nature of the dependence. A coordinate basis transformation T transforms (A, B, C) into ($T^{-1}AT$, $T^{-1}B$, CT) and (F,G) into ($T^{-1}F$, GT). The operators $\mathcal{V}$ and $\mathcal{W}$ are unaltered, while $(\mathcal{V}_A, \mathcal{W}_A) \rightarrow (\mathcal{V}_A T, T^{-1}\mathcal{W}_A)$.

Noting (3.2.10), we conclude that the sensitivity operators transform according to

$$\partial\mathcal{K}_{k,l}/\partial A \rightarrow T^T \partial\mathcal{K}_{k,l}/\partial A\ T^{-T} \tag{3.3.1a}$$

$$\partial\mathcal{K}_{k,l}/\partial B \rightarrow T^T \partial\mathcal{K}_{k,l}/\partial B \tag{3.3.1b}$$

$$\partial\mathcal{K}_{k,l}/\partial C \rightarrow \partial\mathcal{K}_{k,l}/\partial C\ T^{-T} \tag{3.3.1c}$$

Parenthetically, one can note that the corresponding formulae associated with (3.2.9) are not so attractive.

Let us make the definitions

$$J^B{}_{k,l} = \int_0^\tau dt \int_{-\infty}^t (\partial\mathcal{K}_{k,l}(t,s)/\partial B)\ \ (\partial\mathcal{K}_{k,l}(t,s)/\partial B)^T ds \tag{3.3.2a}$$

$$J^C{}_{k,l} = \int_0^\tau dt \int_{-\infty}^t (\partial\mathcal{K}_{k,l}(t,s)/\partial C)^T\ \ (\partial\mathcal{K}_{k,l}(t,s)/\partial C) ds \tag{3.3.2b}$$

Evidently, under the coordinate basis change,

$$J^B{}_{k,l} \rightarrow T^T\ J^B{}_{k,l}\ T \tag{3.3.3a}$$

$$J^C{}_{k,l} \rightarrow T^{-1}\ J^C{}_{k,l}\ T^{-T} \tag{3.3.3b}$$

Make the further definitions

$$J_B = \Sigma_{k,l}\ J^B{}_{k,l} \tag{3.3.4a}$$

$$J_C = \Sigma_{k,l}\ J^C{}_{k,l} \tag{3.3.4b}$$

Then the second and third terms of the measure M_2 are precisely tr J_B and tr J_C. Further, denoting the value of the measure M_2 after coordinate basis transformation by $M_2(T)$ we see that

$$M_2(T) = \Sigma_{k,l}\ \|T^T\ \partial\mathcal{K}_{k,l}/\partial A\ T^{-T}\|_2{}^2 + \mathrm{tr}(J_B P) + \mathrm{tr}(J_C P^{-1}) \tag{3.3.5}$$

where

$$P = T\ T^T \tag{3.3.6}$$

It remains to consider in more details how the first summand in M_2 transforms when there is a coordinate basis change. Notice that

$$\begin{aligned}
&\Sigma_{k,l} \|T^T \partial\mathcal{K}_{k,l}/\partial A\, T^{-T}\|_2{}^2 = \Sigma_{k,l} \int_0^\tau dt \int_{-\infty}^t \|T^T \partial\mathcal{K}_{k,l}(t,s)/\partial A\, T^{-T}\|_F{}^2\, ds \\
&= \Sigma_{k,l} \int_0^\tau dt \int_{-\infty}^t \mathrm{tr}\, [T^{-1}\, (\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T\, T\, T^T\, \partial\mathcal{K}_{k,l}(t,s)/\partial A\, T^{-T}]\, ds \\
&= \Sigma_{k,l}\, \mathrm{tr} \int_0^\tau dt \int_{-\infty}^t \partial\mathcal{K}_{k,l}(t,s)/\partial A\, T^{-T}T^{-1}\, (\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T\;\; T\, T^T\, ds \\
&= \Sigma_{k,l}\, \mathrm{tr} \int_0^\tau dt \int_{-\infty}^t \partial\mathcal{K}_{k,l}(t,s)/\partial A\, P^{-1}\;\; (\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T\;\; P\, ds
\end{aligned} \tag{3.3.7}$$

Hence we can regard M_2 as a function of P:

$$\begin{aligned}
&M_2(P) \\
&= \mathrm{tr}(J_B P) + \mathrm{tr}(J_C P^{-1}) + \Sigma_k \Sigma_l\, \mathrm{tr} \int_0^\tau dt \int_{-\infty}^t \partial\mathcal{K}_{k,l}(t,s)/\partial A P^{-1}\;\; (\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T\, P\, ds
\end{aligned} \tag{3.3.8}$$

The optimal FWL compensator design can thus be formulated as follows:

$$P_{opt} = \arg \min_{P:\, P>0} M_2(P)$$

If a solution exists, then any square root T_{opt} such that $P_{opt} = T_{opt}\, T_{opt}{}^T$ defines an optimal coordinate basis for the controller. We have an expression for the gradient of P:

$$\begin{aligned}
\partial M_2(P)/\partial P = J_B - P^{-1}J_C P^{-1} + \Sigma_k \Sigma_l \int_0^\tau dt \int_{-\infty}^t &[\partial\mathcal{K}_{k,l}(t,s)/\partial A\, P^{-1}\, (\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T \\
&- P^{-1}(\partial\mathcal{K}_{k,l}(t,s)/\partial A)^T\, P\;\; \partial\mathcal{K}_{k,l}(t,s)/\partial A\, P^{-1}]\, ds
\end{aligned} \tag{3.3.10}$$

For evaluation purposes, the following formula is valuable. It is derived using standard properties of the Kronecker product and function vec defined by Neudecker in Neudecker (1969) and proven in a number of textbooks and papers (e.g. Brewer (1978)).

$$\begin{aligned}
&\partial M_2(P)/\partial P = J_B - P^{-1}J_C P^{-1} + \Sigma_k \Sigma_l \mathrm{vec}^{-1}\{\; [\int_0^\tau dt \int_{-\infty}^t (\partial\mathcal{K}_{k,l}(t,s)/\partial A \otimes \partial\mathcal{K}_{k,l}(t,s)/\partial A) ds]\, \mathrm{vec} P^{-1}\} \\
&- \Sigma_k \Sigma_l\, \mathrm{vec}^{-1}\{(P^{-1} \otimes P^{-1})[\int_0^\tau dt \int_{-\infty}^t (\partial\mathcal{K}_{k,l}(t,s)/\partial A \otimes \partial\mathcal{K}_{k,l}(t,s)/\partial A) ds]^T\, \mathrm{vec}\, P\}
\end{aligned} \tag{3.3.11}$$

In the absence of an analytically computable value of P producing zero value of the gradient, a value P_{opt} of P minimizing M_2 could be sought by an iterative algorithm

$$P_{i+1} = P_i - \mu\ \partial\ M_2(P)/\partial P\ |_{P=Pi}, \tag{3.3.12}$$

where μ is a small positive number.

The utility of the gradient algorithm is partly justified by:

Theorem. 3.2
Adopt the same hypothesis as in Theorem 3.1, and let (A, B, C, D) be an initial realization of the controller K(z). Let J_B and J_C be defined by (3.3.2) - (3.3.4), and let $M_2(P)$ define the sensitivity measure (3.3.8) of a controller realization obtained by transforming the initial realization through a nonsingular T, with $P = T\ T^T$. Then there exists a unique $P_{opt} > 0$ which (globally) minimizes $M_2(P)$, which accordingly can be found via an iterative gradient descent algorithm. (There always exist such small positive μ's that guarantee the convergence).

The proof is described in Appendix F.

Once P_{opt} has been found, any square T satisfying $T\ T^T = P_{opt}$ can be selected. This defines T to within right multiplication by an orthogonal matrix, and this additional freedom, present also in all earlier optimal realization problems, can be exploited to force zero or unity entries into parts of A, B or C (see e.g. Li *et al.* (1992)); this has a beneficial practical effect, since obviously a zero or unity multiplication is realizable with no error.

3.4 Evaluation of the Sensitivity Measure and its Gradient

Now, in order to implement the iterative algorithm (3.3.12), we need to calculate the value of the gradient $\partial M_2(P)/\partial P$ at every iteration step. The problem is to calculate the values of

$$\int_0^\tau dt \int_{-\infty}^{t} (\partial\mathcal{K}_{k,l}(t,s)/\partial A \otimes \partial\mathcal{K}_{k,l}(t,s)/\partial A)\, ds,$$

$$\int_0^\tau dt \int_{-\infty}^{t} (\partial\mathcal{K}_{k,l}(t,s)/\partial B)\, (\partial\mathcal{K}_{k,l}(t,s)/\partial B)^T ds \tag{3.4.1}$$

$$\text{and} \quad \int_0^\tau dt \int_{-\infty}^{t} (\partial\mathcal{K}_{k,l}(t,s)/\partial C)^T\, (\partial\mathcal{K}_{k,l}(t,s)/\partial C) ds.$$

The prime concern of this Section is to obtain a numerical procedure for calculating the three values above using standard techniques, i.e. standard software.

Fig. 3.4 depicts the operators $\partial\mathcal{K}_{k,l}/\partial a_{i,j}$, $\partial\mathcal{K}_{k,l}/\partial b_{i,j}$ and $\partial\mathcal{K}_{k,l}/\partial c_{i,j}$. To understand these figures recognize from (3.2.9) that

$$\partial\mathcal{K}_{k,l}/\partial a_{i,j} = (e_k^T\, \mathcal{V}_A\, e_i)\, (e_j^T\, \mathcal{W}_A\, e_l),$$

$$\partial\mathcal{K}_{k,l}/\partial b_{i,j} = (e_k^T\, \mathcal{V}_A\, e_i)\, (e_j^T\, \mathcal{W}\, e_l) \tag{3.4.2}$$

and

$$\partial\mathcal{K}_{k,l}/\partial c_{i,j} = (e_k^T\, \mathcal{V}\, e_i)\, (e_j^T\, \mathcal{W}_A\, e_l).$$

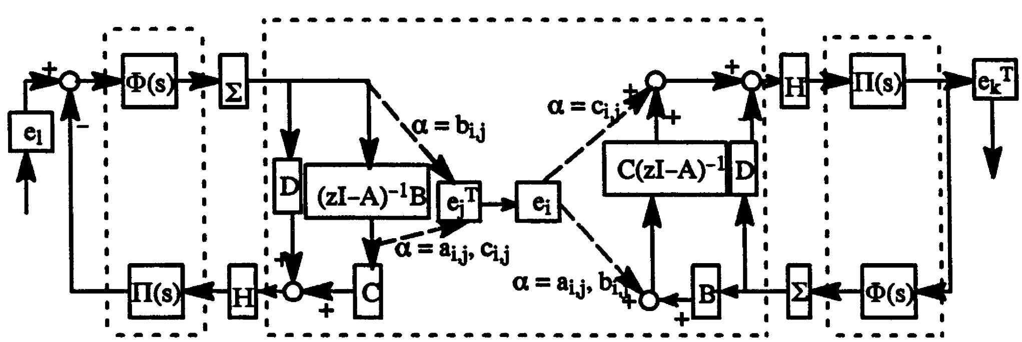


Figure 3.4: The operator $\partial\mathcal{K}_{k,l}/\partial\alpha$ for various α

The structure of this hybrid feedback loop is illustrated in Fig. 3.5, where the form of $\Delta_{d\alpha}(z)$ depends on the particular α. Note the implicit definitions of $\Delta_{d\alpha}(z)$ and $\Omega(s)$.

Because of the mixture of continuous and discrete-time entities, some of the mappings in the hybrid system are operators which have no transfer-function representation and thus the operators $\mathcal{V}_A$ and $\mathcal{W}_A$ do not have transfer function

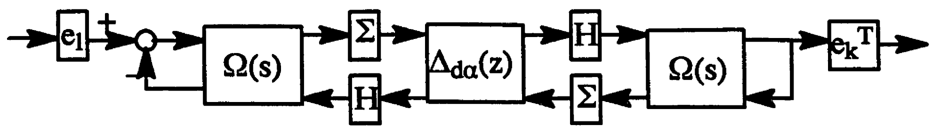


Figure 3.5: Redrawing of Figure 3.4

representations. In order to facilitate the calculations of the L_2 norms, the continuous-time part of the hybrid system is approximated by a discrete-time system with arbitrarily fast sampling. This can be done in a chosen (sensible) frequency range as accurately as desired by hold-input discretization. In this approximation the former hybrid system is replaced by an N-periodic discrete-time system, with the small (fast) sampling time chosen to be a submultiple τ/N of the controller sampling time τ. By lifting the N-periodic control system a time-invariant discrete-time transfer-function representation is obtained. A similar approach has been used for a controller discretization problem in Keller and Anderson (1992), for which it becomes easy to evaluate the norms.

This technique of fast sampling will allow us to approximate the integrals of (3.4.1), taken over one (slow) sampling period τ, by the average of their N sampled values over the period τ for N sufficiently large. To establish the validity of this procedure, we shall show that these sums converge, as $N\to\infty$, to the integrals (3.4.1), using the definition of Riemann integral. This proof of convergence, in turn, requires that the impulse responses of the operators defined in (3.4.2) be continuous and exponentially stable. The following lemma establishes this result. (The proof is straightforward and is omitted).

Lemma 3.1

Under the hypotheses of Theorem 3.1, and the assumption that $\Omega(\infty) = 0$ and that the closed loops depicted in Figs. 3.1 and 3.5 are exponentially stable, (stability of the closed loop in Fig. 3.1 implies stability of the closed loop in Fig. 3.5), the impulse response $h(\cdot,\cdot)$ of the overall system in Fig. 3.5 is continuous for $j\tau < s < (j+1)\tau$, $t>s$ for every integer j, and satisfies $|h(t,s)|\leq\alpha \exp[-\beta(t-s)]$ $\forall\ t > s$, for some positive α and β.

Now let $h_\alpha(t,s)$ again denote the impulse response of any one of the operators $\partial\mathcal{K}_{k,l}/\partial\alpha$ for $\alpha = a_{ij}$, or b_{ij}, or c_{ij}, as depicted in Fig. 3.5. These operators are periodic with period τ. Consider now the system defined in Fig. 3.6, in which $H_{\tau/N}$ and $\Sigma_{\tau/N}$ are, respectively, a hold operator and a sampler operating at the fast rate τ/N. Thus, the system of Fig. 3.6 has discrete time inputs spaced τ/N seconds apart, and a similar output stream. The next lemma expresses its impulse response $h_{d\alpha}(i,j)$ as a function of $h_\alpha(t,s)$ and shows that it is N-periodic. Again, the straightforward proof is omitted.

$$\longrightarrow \boxed{H_{\tau/N}} \longrightarrow \boxed{h_\alpha(t,s)} \longrightarrow \boxed{\Sigma_{\tau/N}} \longrightarrow$$

Figure 3.6: Replacement of periodic continuous-time system by periodic discrete-time system

Lemma 3.2

Let $h_{d\alpha}(i,j)$ denote the impulse response of the system of Fig. 3.6, formally $\Sigma_{\tau/N}\ h_\alpha(t,s)\ H_{\tau/N}$, where $h_\alpha(t,s)$ is identified with $\partial \mathcal{K}_{k,l}(t,s)/\partial\alpha$ for same α. Then

$$h_{d\alpha}(i,j) = \tau/N\ h_\alpha(i\ \tau/N,\ s_{i,j}) \tag{3.4.3}$$

for some $s_{i,j} \in (j\ \tau/N,\ (j+1)\ \tau/N)$ and

$$h_{d\alpha}(i+N,\ j+N) = h_{d\alpha}(i,\ j). \tag{3.4.4}$$

Next, consider a system obtained from that of Fig. 3.6 by blocking N successive inputs and N successive outputs. Thus, if $u_0, u_1, \ldots$ and $y_0, y_1, \ldots$ denote the scalar input and output sequences of the system of Fig. 3.6, with rate (N/τ) per second, $[u_0\ u_1 \ldots u_{N-1}]^T$, $[u_N\ u_{N+1} \ldots u_{2N-1}]^T, \ldots$ and $[y_0\ y_1 \ldots y_{N-1}]^T$, $[y_N\ y_{N+1} \ldots y_{2N-1}]^T, \ldots$ denote the N-vector input and output sequences of the new system, with rate $1/\tau$ per second, 1/N times the rate of the system of Fig. 3.6. Let $\tilde{h}_\alpha(i,j)$ denote the $N \times N$ impulse response of the new system. A moment's thought shows that

$$[\ \tilde{h}_\alpha(i,j)]_{p,q} = h_{d\alpha}(Ni + (p-1),\ Nj + (q-1)) \tag{3.4.5}$$

An immediate consequence of the periodicity of $h_{d\alpha}$ established in Lemma 3.2 above is the fact that the blocked system is stationary

$$\tilde{h}_\alpha(i,j) = \tilde{h}_\alpha(i-j,\ 0).$$

[Hereforth, we shall write $\tilde{h}_\alpha(i-j)$ for $\tilde{h}_\alpha(i-j,\ 0)$]

We remark that given a state-variable description of the various parts of $h_\alpha(t,s)$, it is easy to get such descriptions for $h_{d\alpha}$ and then $\tilde{h}_{\alpha}$. This turns out to be important when it comes to evaluating norms.

The above allows us to evaluate J_B, at least approximately, using time-invariant quantities. We have

Lemma 3.3

With notations as above, the (i-j)-th entry of

$$J^B{}_{k,l} = \int_0^{\tau} dt \int_{-\infty}^{t} (\partial \mathcal{H}_{k,l}(t,s)/\partial B) \;\; (\partial \mathcal{H}_{k,l}(t,s)/\partial B)^T ds$$

is given by

$$(J^B{}_{k,l})_{i,j} = \lim_{N \to \infty} \sum_{m=1}^{L} \sum_{s=0}^{\infty} \operatorname{tr} [\tilde{h}_{B\,im}(s)\, \tilde{h}_{B\,jm}{}^T(s)],$$

where $\tilde{h}_{B\,im}(s)$ denotes $\tilde{h}_{\alpha}(s)$ for $\alpha = b_{im}$, the (i-m)-th entry of B.

The proof is contained in Appendix G.

Quantities such as $(J^B{}_{k,l})_{ij}$ above are like Gramians and are comparatively easy to evaluate. References Åström and Wittenmark (1990) and Jury (1958) discuss such quantities, and offer several methods, especially when i = j. Let us point out that the identity $\alpha\beta = [(\alpha+\beta)^2 - (\alpha-\beta)^2]/4$ offers one device to cope with $i \neq j$, if formulae are only available for evaluating sums of squares. Let us also note how simple it is to use linear matrix or Lyapunov equations for evaluation of infinite sums involving a product of two different stable impulse responses. Thus if

$$\alpha_1(k) = h_1{}^T F_1{}^{k-1} g_1 \qquad \alpha_2(k) = h_2{}^T F_2{}^{k-1} g_2$$

then, assuming $|\lambda_i(F_j)| < 1$ for all i and j = 1,2

$$\sum_{k=1}^{\infty} \alpha_1(k)\, \alpha_2(k) = h_1{}^T X\, h_2$$

where X solves

$$X - F_1 X F_2{}^T = g_1 g_2{}^T.$$

Similarly, it follows that the (i-j)-th entry of

$$J^{C}_{k,l} = \int_0^{\tau} dt \int_{-\infty}^{t} (\partial \mathcal{H}_{k,l}(t,s)/\partial C)^T \quad (\partial \mathcal{H}_{k,l}(t,s)/\partial C) ds$$

is given by

$$(J^{C}_{k,l})_{i,j} = \lim_{N\to\infty} \sum_{m=1}^{L} \sum_{s=0}^{\infty} \mathrm{tr}\, [\tilde{h}_{C\, mi}^T(s)\, \tilde{h}_{C\, mj}(s)].$$

The calculations are identical to those for J_B. It remains to evaluate the first integral of (3.4.1).

A typical entry of the matrix

$$\int_0^{\tau} dt \int_{-\infty}^{t} (\partial \mathcal{H}_{k,l}(t,s)/\partial A \otimes \partial \mathcal{H}_{k,l}(t,s)/\partial A)\, ds$$

is given by

$$\int_0^{\tau} dt \int_{-\infty}^{t} h_{A\, mn}(t,s)\, h_{A\, pq}(t,s)\, ds$$

and similar arguments show that this quantity is obtainable as

$$\lim_{N\to\infty} \sum_{s=0}^{\infty} \mathrm{tr}\, [\tilde{h}_{A\, mn}(s)\, \tilde{h}_{A\, pq}(s)].$$

It remains to explain how to evaluate $\tilde{h}_{\alpha}$ for $\alpha = a_{ij}$, or b_{ij}, or c_{ij}. For this purpose Figs. 3.7a, 3.7b and 3.7c are helpful, and illustrate a certain computational simplification. Fig. 3.7a is Fig. 3.6 redrawn, while in Fig. 3.7b, which is equivalent to Fig. 3.7a, each sampler Σ within the hybrid system (which selects a sample every τ seconds) is replaced by a sampler $\Sigma\tau/_N$, selecting a sample every τ/N seconds, followed by a decimator, which passes through every N–th input. Each hold of duration τ is replaced by a repeater (which repeats a signal presented at a given time with the same value τ/N, $2\tau/N$, . . . , $(N-1)\tau/N$ seconds later) and a hold of duration τ/N.

The dashed line encloses a fast discrete–time system (which can be lifted or

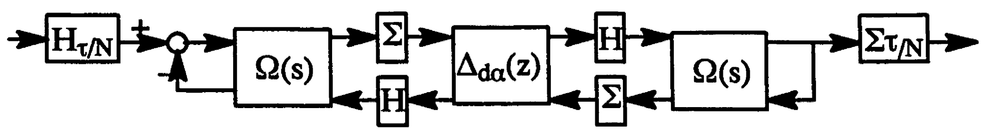


Figure 3.7a: Redrawing of Figure 3.6

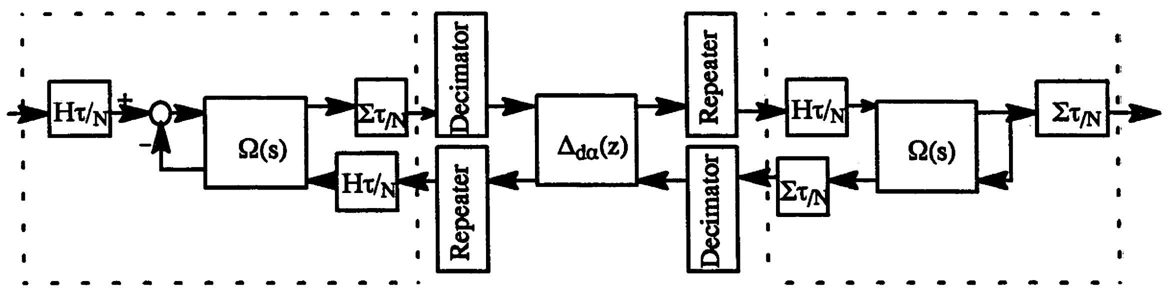


Figure 3.7b: Replacement of Figure 3.7a

blocked), and the decimator and repeater serve to connect the discrete-time blocks with different sampling rates. Lifting produces the arrangement of Fig. 3.7c, where the input and output values of $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are obtained by assembling into the one vector N successive values of the input and output of each of the blocks in dashed line in the set-up of Fig. 3.7b.

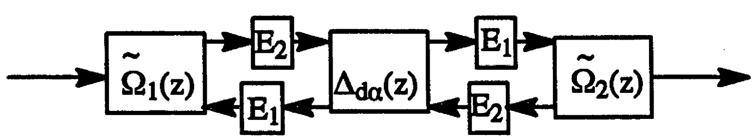


Figure 3.7c: Development of scheme of Figure 3.6
and the lifted system

The lifted set-up has $E_1=(I\ I\ ...\ I)^T$ and $E_2=(I\ 0\ 0\ ...\ 0)$ and is single-rate, with period τ, and possesses a transfer function description. The impulse response is $\tilde{h}_\alpha(i)$. Since it is the same set-up as Fig. 3.7a (apart from the way inputs and outputs are presented), it is no surprise that norms of equivalent input-output entities are the same. (Equivalent means after allowing for the different assembling of inputs/outputs). The structure of Fig. 3.7c demonstrates that a state-variable realization for $\tilde{h}_\alpha$ comes from assembling realizations for $\tilde{\Omega}_1$, $\tilde{\Omega}_2$ (which are independent of α) and of $\Delta_{d\alpha}$.

Evidently, the problem of calculating $\partial M_2(P)/\partial P$ is reduced to the problem of calculating infinite sums which is much simpler, and involves standard computational techniques implemented on most software packages. That makes the iterative algorithm (3.3.12) easy to realize.

To summarize, the main steps of the optimal controller realization procedure are as follows, assuming that K(z) is the ideal (infinite wordlength) controller.

Step 1: Compute an arbitrary initial realization (A, B, C, D) of K(z) and initialize the iterative algorithm (3.3.12) with $P_0 = I$.

Step 2: For k,l = 1, 2, . . . , L, approximate $J^B_{k,l}$ of (3.3.2a) as described in Steps 2.1–2.3, with N sufficiently large and $(J^B_{k,l})_{ij}$ denoting the (i–j)–th entry of $J^B_{k,l}$.
Note that $J^B_{k,l} \in \mathbb{R}^{R\times R}$.

Step 2.1: For each $\alpha = b_{im}$, m=1, 2, . . . , L and each $\alpha = b_{jm}$, m=1, 2, . . . , L, let $h_\alpha(t,s)$ be the impulse response defined by Fig. 3.4. Obtain a state–variable description of $h_\alpha(t,s)$ using state–variable descriptions of the different blocks of Fig. 3.4. Compute the (fast sampled) ZOH approximation of this system (see Fig. 3.6) whose impulse response $h_{d\alpha}(i,j)$ is defined by (3.4.1):

$$h_{d\,\alpha}(i,j)=\tau/N\ h_\alpha(i\tau/N, j\tau/N), \text{ say.}$$

Compute the corresponding N × N blocked system of Fig. 3.7c, whose impulse response is

$\tilde{h}_\alpha(i,j) = \tilde{h}_\alpha(i-j)$, given by (3.4.3):

$$[\,\tilde{h}_\alpha(i-j)]_{p,q} = h_{d\alpha}(Ni + p - 1, Nj + q - 1) \quad p, q = 1, 2, \ldots, N.$$

Step 2.2: Approximate $(J^B_{k,l})_{ij}$ by

$$(J^B_{k,l})_{i,j} \approx \sum_{m=1}^{L} \sum_{s=0}^{\infty} \operatorname{tr}\,[\tilde{h}_{B\,im}(s)\,\tilde{h}_{B\,jm}{}^T(s)],$$

where $\tilde{h}_{B\,im}(s) = \tilde{h}_\alpha(s)$ for $\alpha = b_{im}$, using the Lyapunov equation procedure suggested in Section 3.4.

Step 2.3: Compute $J_B = \Sigma_{k,l}\, J^B_{k,l}$.

Step 3: Compute (or rather approximate) J_C using a procedure entirely dual to that for J_B.

Step 4: Computation of

$$\int_0^\tau dt \int_{-\infty}^{t} (\partial \mathcal{K}_{k,l}(t,s)/\partial A \otimes \partial \mathcal{K}_{k,l}(t,s)/\partial A)\, ds.$$

The approximation of this integral is performed as follows.

Step 4.1: For each $\alpha = a_{mn}$, m, n = 1, 2, . . . , R let $h_\alpha(t,s)$ be the impulse response defined by Fig. 3.4. By fast sampling and blocking, compute a state-space realization of the discrete time impulse response of the corresponding fast-sampled and blocked system as in Step 2.2. Denote by $\tilde{h}_{A\,mn}(s)$ the $N \times N$ impulse response matrix of the corresponding blocked system.

Step 4.2: The element of

$$\int_0^\tau dt \int_{-\infty}^{t} (\partial \mathcal{K}_{k,l}(t,s)/\partial A \otimes \partial \mathcal{K}_{k,l}(t,s)/\partial A)\, ds$$

corresponding to the product of the (m–n)–th entry of the first matrix with the (p–q)–th entry of the second matrix is then approximated by

$$\sum_{s=0}^{\infty} \mathrm{tr}\, [\tilde{h}_{A\,mn}(s)\, \tilde{h}_{A\,pq}(s)].$$

To compute this infinite sum, the state–space realizations of $\tilde{h}_{A\,mn}(s)$ and $\tilde{h}_{A\,pq}(s)$ are used in combination with the Lyapunov equation technique of Section 3.4.

Step 5: Collect the results of Steps 2, 3 and 4 in equation (3.3.11) to compute the gradient $\partial M_2(P)/\partial P$ and update P_i using the iterative algorithm (3.3.12).

Step 6: Upon convergence of (3.3.12) to P_{opt}, compute any square root T_{opt} such that $P_{opt} = T_{opt}\, T_{opt}^T$ and apply the similarity transformation T_{opt} to the initial realization (A, B, C, D) of the compensator K(z) to obtain an optimal realization (A_{opt}, B_{opt}, C_{opt}, D_{opt}). Optionally, introduce a further orthogonal transformation to force zero entries into A_{opt}, B_{opt} and/or C_{opt} if desired, see e.g. Li *et al.* (1992).

3.5 Numerical Examples

We now present two numerical examples to confirm our theoretical results. The first example is a simple one with the one-state controller, has no applied interest, but allows us to get better understanding of the system's properties and behaviour. The second example has been used in Ackermann (1985).

3.5.1 Example 3.1

The plant to be controlled is given by its transfer function

$\Pi(s) = (s + .9531) / (s - .0953)$.

The desirable control strategy is to control this plant in such way that the closed loop has the following transfer function:

$X(s) = .8318 / (s + .6931)$.

The controller to be used with a sampler and a zero-order hold with the sampling period $\tau = 1$ has the following transfer function:

$K(z) = .6 / z$.

Let us consider two realizations of the controller

$K(z) = b\ c / (z-a) + d$.

The first one is with $a = d = 0$, $b = .006$, $c = 100$ and the second one is with
$a = d = 0$,
$b_{opt} = c_{opt} = \sqrt{.6}$.

The second realization is the optimal one. For this one-state controller both the L_2 measure minimization (using fast-sampling and blocking) described in this Chapter and optimization without fast-sampling give the same optimal realization. By optimization without fast sampling, we mean optimization using a discrete-time representation of the plant obtained with the same sampling interval as for the controller. (Equivalently, it is like fast sampling and blocking with N=1!)

When we implement our two realizations of the controller with FWL giving roundoff with two decimal places after the decimal point, we obtain
$K(z) = 1 / z$ for the first realization and $K_{opt}(z) = .59 / z$.
These controllers give the closed loops
$X(z) = 1 / (z - 0.1)$ and $X_{opt}(z) = 0.59 / (z - 0.51)$.

The frequency responses of these closed loops together with the frequency response of the ideal (realized with infinite precision) closed loop are depicted in Fig. 3.8. Here, by frequency response of a sampled system we understand the frequency response of the discrete system obtained by interconnecting the discrete controller and a discrete model of the plant, obtained from the zero order hold equivalent of the continuous model. Though it is not plotted, the closed-loop response of X(s) is very close to the "ideal frequency response" of Fig. 3.8.

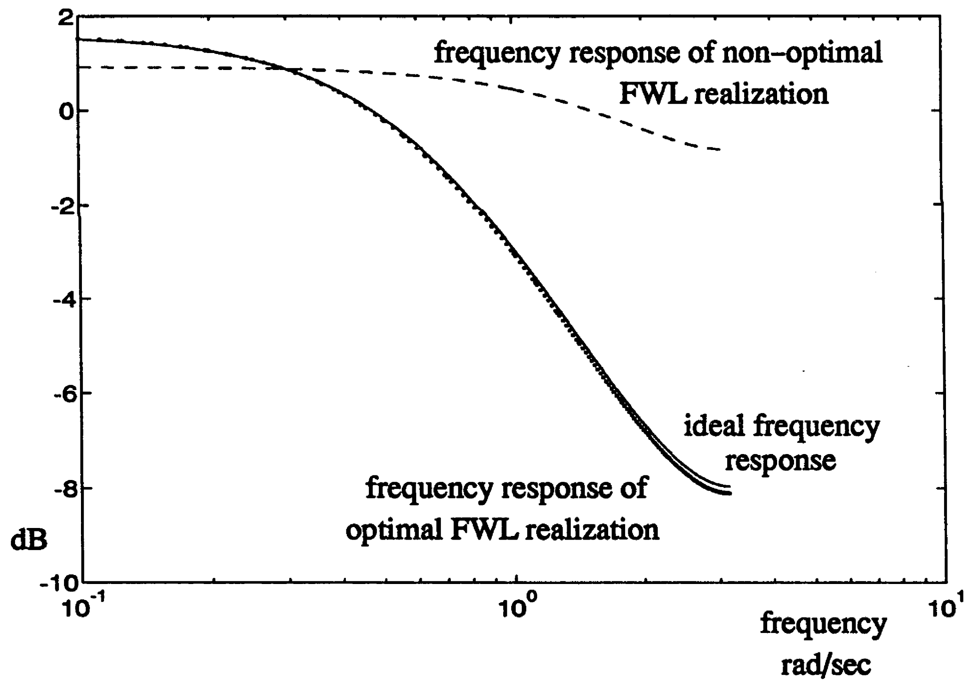


Figure 3.8: Closed-loop frequency responses (Example 3.1)

Obviously, the optimal FWL realization gives incomparably better approximation of the desired ideal loop.

3.5.2 *Example 3.2*, see Ackermann (1985), p.239.

The plant to be controlled is given by its transfer function

$\Pi(s) = 1 / (s + 1)$.

The controller whose output is the input of a zero-order hold and whose input is sampled with the period $\tau = 2$ has the following transfer function:

$K(z) = 1 / (z - 1)^2$.

Consider the realization of the controller given by

$$A = \begin{pmatrix} 4.5107 & -.7742 \\ 15.918 & -2.5107 \end{pmatrix}, \quad B = \begin{pmatrix} 5.9619 \\ 21.3939 \end{pmatrix},$$

$$C = (-.8688 \quad .2421), \quad D = 0.$$

as an initial one.

FWL optimization, not employing fast-sampling/blocking (or, equivalently, employing them for N=1), but using simply a discrete-time representation of the plant and the closed loop, thus neglecting intersample behaviour, gives the following realization:

$$A = \begin{pmatrix} .7998 & -.0404 \\ .9981 & 1.1998 \end{pmatrix}, \quad B = \begin{pmatrix} .8859 \\ 1.1023 \end{pmatrix},$$

$$C = (-.9031 \quad .7231), \quad D = 0.$$

Minimization of the M_2 measure for the pre-fastsampled and then blocked system gives the optimal realization of the controller given by

$$A = \begin{pmatrix} .6556 & -.2223 \\ .5335 & 1.3444 \end{pmatrix}, \quad B = \begin{pmatrix} 1.0507 \\ .8551 \end{pmatrix},$$

$$C = (-.624 \quad .7668), \quad D = 0.$$

By using an orthogonal transformation of the state basis, we do not change the formally defined sensitivity and thus optimality. However, by bringing A to Schur form (Li *et al.*, 1992) we can incorporate a zero into the matrix to make computations even more precise, as zero has an infinitely precise computer representation.

The frequency responses of the closed loops corresponding to these two optimal realizations of the closed loop (obtained by different procedures) and the frequency response of the closed loop with the controller given by an initial non-optimal realization implemented with one decimal place after the decimal point roundoff and the frequency response of the ideal (realized with infinite precision) closed loop are represented in Fig. 3.9.

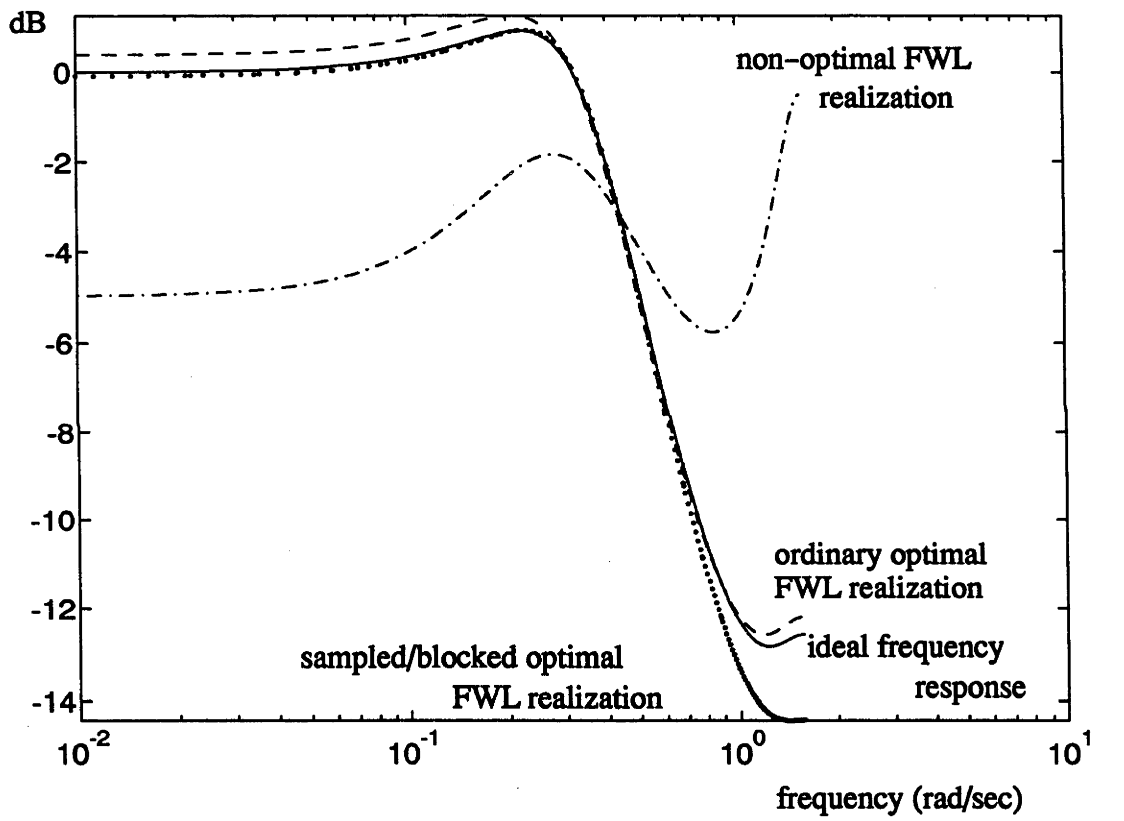


Figure 3.9: Closed-loop frequency responses (Example 3.2)

The superiority within the passband of the closed-loop system is clearly seen of the optimal sensitivity realization, obtained by fast sampling and blocking of the system, over the optimal sensitivity realization neglecting intersample behaviour of the system. Also, both optimal sensitivity realizations are incomparably superior to the initial non-optimal realization.

3.6 Conclusions

The proposed method obtains the FWL-optimal realization of a discrete-time controller which is used in a closed loop with a continuous-time plant, a sampler, a zero-order hold and an antialiasing filter and which minimize a sensitivity index. This optimal realization is based on complete information describing the closed-loop system's behaviour, not only at the sampling instances but in inter-sample periods as well. The existence and uniqueness of this optimal realization (to within an orthogonal coordinate basis transformation) have been established and a recursive algorithm converging to the realization has been given.

The theoretical results have been confirmed by two numerical examples which illustrate the feasibility and efficiency of the proposed method and the advantage of taking into account inter-sample behaviour of a closed-loop system.

Why sir, there is every possibility
that you will soon be able to tax it!

Michael Faraday to Gladstone, when asked
about the usefulness of electricity;
in W.E.H.Lecky *Democracy and Liberty*

Chapter 4

Sampled–Data Controller Reduction

4.1 Introduction

Model reduction by means of balanced realizations and Hankel–norm approximations has been studied by Moore (1981) and Glover (1984). Enns introduced frequency weighting to balanced realizations [1984a] and applied this approach for maintaining closed–loop stability [1984b]. Anderson and colleagues have developed a frequency–weighted Hankel–norm technique for controller reduction (Anderson and Liu, 1989, Latham and Anderson, 1985). None of this earlier work explicitly treated sampled–data systems.

In this Chapter, our objective is to apply a balanced realization controller order reduction method to sampled–data closed–loop systems to preserve the closed–loop behaviour.

An outline of this Chapter is as follows. In Section 4.2 we apply the previously introduced fast sampling and lifting scheme to the sampled–data closed–loop system. A time-invariant system results. In Section 4.3 we obtain the weighting functions for preserving the closed-loop transfer function and actually reduce the controller by the weighted balanced truncation method. A practical example to confirm the approach is given in Section 4.4, followed by some concluding remarks in Section 4.5.

4.2 Fast Sampling and Lifting

The purpose of this section is to apply the introduced in Chapter 2 fast sampling and lifting operation to the sampled-data system of Fig. 1.1, i.e. to replace the periodically time-varying sampled-data system (with continuous-time input and output) by a time-invariant system.

To do so one should obtain a discrete-time approximation of the system by sampling and then lift the system as has been described in Khargonekar *et al.* (1985). The sampling interval is τ/N, where τ is the controller sampling time. The sampled system is a multi-rate N-periodic discrete-time system. Lifting involves passing from an N-periodic linear $p \times m$ discrete-time sampled system to an equivalent $pN \times mN$ discrete-time linear time-invariant system. Observe that the equivalence is an isomorphism of the systems in the sense that both essential algebraic and analytic properties of the systems are preserved. In particular, the lifted system is stable if and only if the N-periodic system is stable, and in this case the operator norms (associated with regarding the system as an operator mapping square-summable input to square-summable output) are equal.

In order to take into account intersample behaviour of the system in Fig. 1.1, we introduce fast sampling of the system with the sampling time τ/N chosen to be a submultiple N of the controller sampling time τ, see Fig. 4.1.

Normally, τ/N is chosen to be smaller than the fastest significant time constant in the scheme of Fig. 1.1, e.g. the inverse of $5 \times$ closed-loop bandwidth.

It was shown in Chapter 2 that the performance of the set-up of Fig. 4.1 will mimic that of the scheme of Fig. 1.1. We shall show how a reduced order controller can be obtained for the scheme of Fig. 4.1. If it is satisfactory, in the sense of causing very little error in the closed-loop behaviour of Fig. 4.1,

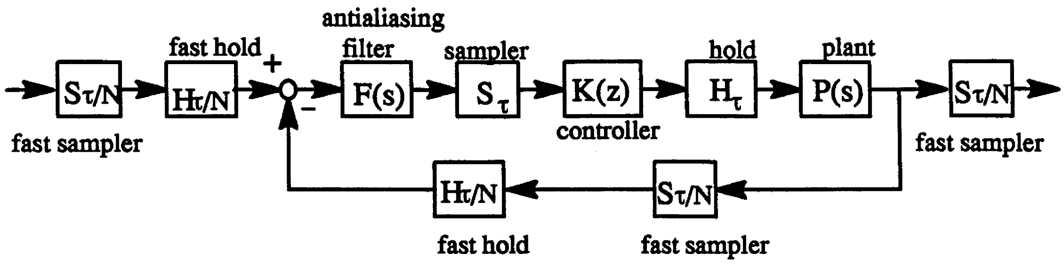


Figure 4.1: The fast-sampled closed-loop system

it will have this property for the scheme of Fig. 1.1 also.

To obtain a time-invariant system, so that standard reduction procedures can be applied, we lift the system in Fig. 4.1. Given state-space realizations of the plant P and antialiasing filter F as

$$P(s)=C_p(sI-A_p)^{-1}B_p+D_p \tag{4.2.1a}$$

$$F(s)=C_f(sI-A_f)^{-1}B_f \tag{4.2.1b}$$

the state-space realizations of the mN-input, pN-output lifted plant $\mathcal{P}$ and the pN-input, pN-output filter $\mathcal{F}$ can be written in the form

$$\mathcal{P}(z)=\mathcal{C}_p(zI-\mathcal{A}_p)^{-1}\,\mathcal{B}_p+\mathcal{D}_p \tag{4.2.2a}$$

$$\mathcal{F}(z)=\mathcal{C}_f(zI-\mathcal{A}_f)^{-1}\,\mathcal{B}_f+\mathcal{D}_f \tag{4.2.2b}$$

where

$$\mathcal{A}_p=a_p^N,\quad \mathcal{B}_p=[a_p^{N-1}b_p \;\; \ldots \;\; a_pb_p \;\; b_p],$$

$$\mathcal{A}_f=a_f^N,\quad \mathcal{B}_f=[a_f^{N-1}b_f \;\; \ldots \;\; a_fb_f \;\; b_f],$$

$$\mathcal{C}_p=[C_p^T \;\; a_p^TC_p^T \;\; \ldots \;\; (a_p^T)^{N-1}C_p^T]^T,$$

$$\mathcal{C}_f=[C_f^T \;\; a_f^TC_f^T \;\; \ldots \;\; (a_f^T)^{N-1}C_f^T]^T,$$

$$\mathcal{D}_p=\begin{pmatrix} D_p & 0 & \ldots & 0 \\ C_pb_p & D_p & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 \\ C_pa_p^{N-2}b_p & C_pa_p^{N-3}b_p & \ldots & D_p \end{pmatrix}$$

$$\mathcal{D}_f = \begin{pmatrix} 0 & 0 & \dots & 0 \\ C_f b_f & 0 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ C_f a_f^{N-2} b_f & C_f a_f^{N-3} b_f & \dots & 0 \end{pmatrix}$$

$a_p = \exp(A_p \tau/N)$, $a_f = \exp(A_f \tau/N)$,

$$b_p = \int_0^{\tau/N} \exp(A_p t) dt\, B_p, \qquad b_f = \int_0^{\tau/N} \exp(A_f t) dt\, B_f$$

(The realizations (4.2.2) are minimal if (4.2.1) are minimal, for almost all choices of τ.)

The lifted controller $\mathcal{K}$ can be written as

$$\mathcal{K}(z) = E_1\, K(z)\, E_2$$

where

$$E_1 = (I_m \;\; I_m \;\; \dots \;\; I_m)^T \;\; \in \;\; \mathbb{R}_{mN\times m},$$

$$E_2 = (I_p \;\; 0_p \;\; 0_p \;\; \dots \;\; 0_p) \in \mathbb{R}_{p\times pN},$$

I_n – $n \times n$ identity matrix, 0_p – $p \times p$ zero matrix.

E_2 in the formula corresponds to a slow (every τ seconds) sampler, which passes through only the first element of an input vector and is in the off mode when the following N–1 elements of the input vector arrive. E_1 corresponds to a τ-second zero-order hold.

Introducing $\overline{P}$ and $\overline{F}$ as

$$\overline{P} = \mathcal{P}\, E_1$$

and

$$\overline{F} = E_2\, \mathcal{F},$$

the periodically time-varying sampled-data system in Fig. 1.1 can be replaced by a time invariant system in Fig. 4.2.

Associated with the closed-loop linear periodically time-varying operator T of Fig. 1.1 is the corresponding linear time invariant operator $\mathcal{T}$ of Fig. 4.2:

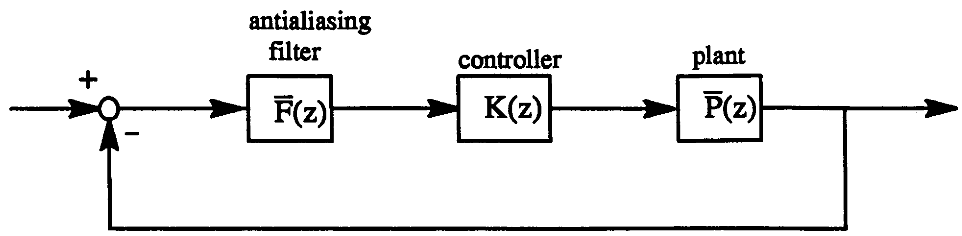


Figure 4.2: The lifted closed-loop system

$$\mathcal{T} = \overline{P}\, K\, \overline{F}\, (I + \overline{P}\, K\, \overline{F})^{-1}. \tag{4.2.3}$$

Controller reduction will be performed with $\mathcal{T}$ in mind, knowing that good controller reduction for $\mathcal{T}$ provides good controller reduction for T.

4.3 Controller Reduction

In this section we will reduce the order of the controller K of the setup in Fig. 4.2 using the approach given in Anderson and Liu (1989). The major aim of this reduction is preservation of the closed-loop transfer function. This means that the error in approximation of the controller K by the reduced order controller K_r is measured by

$$\| W(z)\, [K(z)-K_r(z)]\, V(z) \|_\infty \tag{4.3.1}$$

where weights W and V are dictated by the requirement to preserve (as far as possible) the closed-loop transfer function. In minimizing the error, they cause the approximation process for K to be more accurate at certain frequencies. We shall now determine these weights. Denote by $\mathcal{T}_r$ the transfer function of the closed loop of Fig. 4.2 with the reduced-order controller K_r, and consider the difference

$$\mathcal{T} - \mathcal{T}_r = \overline{P}\, K\, \overline{F}\, (I + \overline{P}\, K\, \overline{F})^{-1} - \overline{P}\, K_r\, \overline{F}\, (I + \overline{P}\, K_r\, \overline{F})^{-1} \tag{4.3.2}$$

To a first order approximation in $K - K_r$, there holds

$$\mathcal{T} - \mathcal{T}_r \approx (I + \overline{P}\,K\,\overline{F})^{-1}\,\overline{P}\,(K - K_r)\,\overline{F}\,(I + \overline{P}\,K\,\overline{F})^{-1} \tag{4.3.3}$$

This suggests the choice of weighting functions

$$W(z) = (I + \overline{P}\,K\,\overline{F})^{-1}\,\overline{P} \tag{4.3.4a}$$

$$V(z) = \overline{F}\,(I + \overline{P}\,K\,\overline{F})^{-1} \tag{4.3.4b}$$

and the minimization problem

$$\min \| W\,(K - K_r)\,V \|_\infty \tag{4.3.5}$$

To solve this problem, at least in an approximate way, we suggest the frequency–weighted balanced truncation technique (Anderson and Liu, 1989, Enns, 1984b) be applied to the stable part of K. (The unstable part of K is copied into K_r). We shall now briefly review this technique. Consider asymptotically stable frequency–weighting functions and associated minimal state–variable realizations

$$W(z) = C_w(zI - A_w)^{-1}\,B_w + D_w$$

and

$$V(z) = C_v(zI - A_v)^{-1}\,B_v + D_v$$

(W and V are stable when the closed loop $\mathcal{T}$ is stable). The basic idea is to change the gramians to reflect the introduction of the frequency weighting. The frequency–weighted transfer function W(z)K(z)V(z) has a representation with the following state–space matrices:

$$\overline{A} = \begin{pmatrix} A_W & B_WC & B_WDC_V \\ 0 & A & BC_V \\ 0 & 0 & A_V \end{pmatrix} \quad \overline{B} = \begin{pmatrix} B_WDD_V \\ BD_V \\ B_V \end{pmatrix}$$

$$\overline{C} = (C_W \quad D_WC \quad D_WDC_V).$$

[Replace K(z) by its stable part in an additive decomposition, if K(z) is not stable].

$$\text{Let } \overline{U} = \begin{pmatrix} U_W & U_{12} & U_{13} \\ U_{12}^T & U & U_{23} \\ U_{13}^T & U_{23}^T & U_V \end{pmatrix} \quad \text{and } \overline{Y} = \begin{pmatrix} Y_W & Y_{12} & Y_{13} \\ Y_{12}^T & Y & Y_{23} \\ Y_{13}^T & Y_{23}^T & Y_V \end{pmatrix}$$

be the solutions of the following Lyapunov equations:

$$\overline{A}\,\overline{U}\,\overline{A^T} + \overline{B}\,\overline{B^T} = \overline{U} \tag{4.3.6a}$$

$$\overline{A^T}\,\overline{Y}\,\overline{A} + \overline{C^T}\,\overline{C} = \overline{Y} \tag{4.3.6b}$$

Now, U and Y can be regarded as the frequency-weighted controllability and observability gramians for the original controller K(z) (or its stable part).

Consider a coordinate basis change to {A, B, C} which makes
$U_{new} = Y_{new} = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_n)$, $\mu_i \geq \mu_{i+1}$, $i=1,2, \ldots, n-1$.
This new realization {A, B, C} is called a frequency-weighted balanced realization.

Now, the controller reduction is achieved by eliminating the rows and columns of A, B and C corresponding to smallest $(\mu_{r+1}, \mu_{r+2}, \ldots, \mu_n)$ in $U_{new} = Y_{new}$. This yields $K_r(z)$ (or its stable part).

A detailed, computer oriented description of this weighted balanced truncation algorithm is given in Integrated Systems Inc. (1991).

This frequency-weighted balanced truncation technique allows one to reduce the controller K(z) preserving as much as possible the closed-loop transfer function $\mathcal{T}$. It will be shown in Chapter 6 that unlike in the non-weighted or single-side weighted case, a stable K(z) may not yield a stable $K_r(z)$. A frequency error bound for the frequency-weighted controller order reduction when stability is preserved will be derived in the following Chapter.

4.4 Example

We now present a practical example to confirm the applicability of the approach. This example has been studied in Anderson and Moore (1989) and Chongsrid and Hara (1994).

The system considered in this example comprises four spinning disks. The disks are connected by a flexible rod, a motor applies torque to the third disk, and the angular displacement of the first disk is the variable of interest. The state-space matrices of the eight state plant are given by

$$A_p=\text{diag}\left\{\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \begin{bmatrix}-0.015 & 0.765\\ -0.765 & -0.015\end{bmatrix}, \begin{bmatrix}-0.028 & 1.410\\ -1.410 & -0.028\end{bmatrix}, \begin{bmatrix}-0.04 & 1.85\\ -1.85 & -0.04\end{bmatrix}\right\}$$

$$B_p^T = [0.026\ \ -0.251\ \ 0.033\ \ -0.886\ \ -4.017\ \ 0.145\ \ 3.604\ \ 0.280],$$

$$C_p = [-0.996\ \ -0.105\ \ 0.261\ \ 0.009\ \ -0.001\ \ -0.043\ \ 0.002\ \ -0.026],\quad D_p=0.$$

Design of the LQG controller was described in Anderson and Moore (1989) and results in the continuous-time controller with the following state-space representation:

$$A_c = \begin{pmatrix}
-0.4077 & 0.9741 & 0.1073 & 0.0131 & 0.0023 & -0.0186 & -0.0003 & -0.0098\\
-0.0977 & -0.1750 & 0.0215 & -0.0896 & -0.0260 & 0.0057 & 0.0109 & -0.0105\\
0.0011 & 0.0218 & -0.0148 & 0.7769 & 0.0034 & -0.0013 & -0.0014 & 0.0011\\
-0.0361 & -0.5853 & -0.7701 & -0.3341 & -0.0915 & 0.0334 & 0.0378 & -0.0290\\
-0.1716 & -2.6546 & -0.0210 & -1.4467 & -0.4428 & 1.5611 & 0.1715 & -0.1318\\
-0.0020 & 0.0950 & 0.0029 & 0.0523 & -1.3950 & -0.0338 & -0.0062 & 0.0045\\
0.1607 & 2.3824 & 0.0170 & 1.2979 & 0.3721 & -0.1353 & -0.1938 & 1.9685\\
-0.0006 & 0.1837 & 0.0048 & 0.1010 & 0.0289 & -0.0111 & -1.8619 & -0.0311
\end{pmatrix}$$

$$B_c^T = [-0.4105\ \ -0.0868\ \ -0.0004\ \ 0.0036\ \ 0.0081\ \ -0.0085\ \ -0.0004\ \ -0.0132],$$

$$C_c = [-0.0447\ \ -0.6611\ \ -0.0047\ \ -0.3601\ \ -0.1033\ \ 0.0375\ \ 0.0427\ \ -0.0329],$$

$$D_c = 0.$$

The controller is open-loop stable.

A discrete-time controller with sampling time τ=0.1 sec. is obtained, by finding the zero-order hold equivalent of the open-loop controller. [This procedure, although it appears satisfactory in this case, can be subject to criticism on the grounds that it does not seek directly to have the continuous-time closed-loop closely approximated by the sampled-data closed loop as in Keller and Anderson (1991, 1992).] The discrete-time controller can be described by the following state-space form:

$$
= \begin{pmatrix}
0.9596 & 0.0945 & 0.0106 & 0.0012 & 0.0002 & -0.0018 & 0.0001 & -0.0010 \\
-0.0094 & 0.9829 & 0.0024 & -0.0084 & -0.0025 & 0.0004 & 0.0011 & -0.0009 \\
-0.0000 & -0.0001 & 0.9956 & 0.0763 & -0.0000 & 0.0000 & 0.0000 & -0.0000 \\
-0.0031 & -0.0556 & -0.0757 & 0.9653 & -0.0089 & 0.0025 & 0.0038 & -0.0024 \\
-0.0148 & -0.2506 & 0.0030 & -0.1361 & 0.9474 & 0.1516 & 0.0172 & -0.0108 \\
0.0008 & 0.0270 & 0.0000 & 0.0147 & -0.1358 & 0.9860 & -0.0018 & 0.0012 \\
0.0139 & 0.2265 & -0.0028 & 0.1229 & 0.0360 & -0.0103 & 0.9635 & 0.1930 \\
-0.0015 & -0.0039 & 0.0002 & -0.0020 & -0.0006 & 0.0002 & -0.1830 & 0.9788
\end{pmatrix}
$$

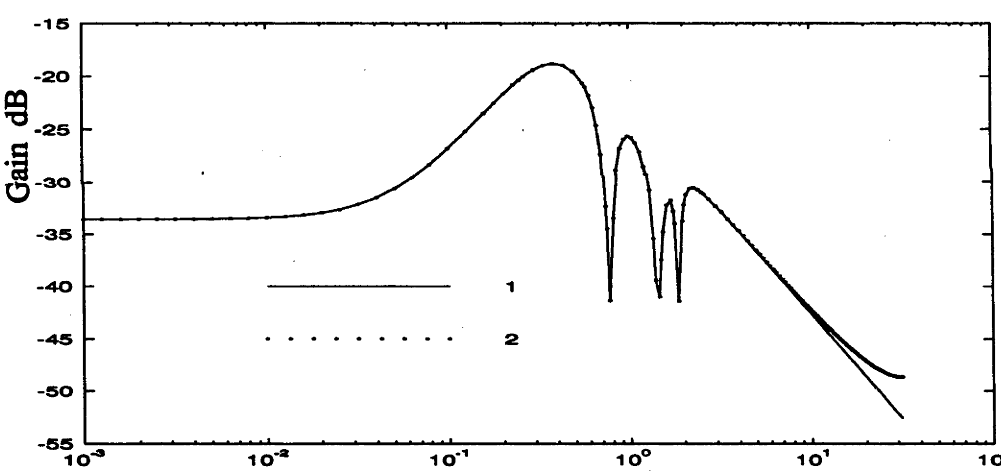
with "A" printed above the matrix:

A

$B^T = [-0.0406\ -0.0084\ -0.0000\ 0.0007\ 0.0021\ -0.0010\ -0.0014\ -0.0013]$,

$C = C_c$, $D = 0$.

The frequency responses of the continuous controller and its sampled version are depicted in Fig. 4.3.



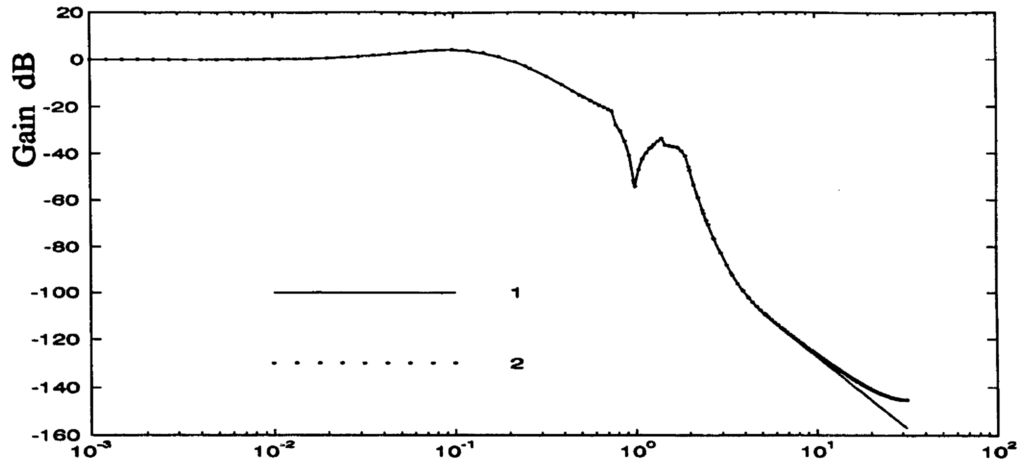
1 continuous controller

2 sampled controller (plotted up to half the sampling frequency)

Figure 4.3: Frequency responses of continuous and sampled controllers

The antialiasing filter has transfer function $F(s) = 5/(s+5)$.

The frequency responses of the initial continuous and sampled closed loops are shown in Fig. 4.4. Here, by frequency response of a sampled system we mean the frequency response of the discrete system obtained using zero-order hold discrete-time equivalents of both the plant and the antialiasing filter.



1 continuous closed loop

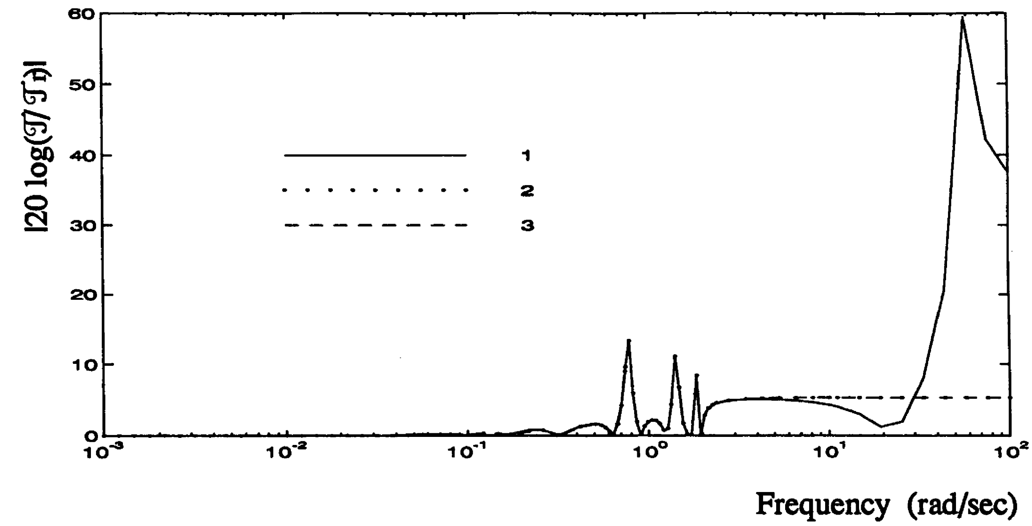
2 sampled closed loop (plotted up to half the sampling frequency)

Figure 4.4: Frequency responses of continuous and sampled closed loops

The controller order was reduced using the fast-sampling and lifting technique with N=3 and N=10 (sampling period τ/N). In both cases, frequency weighting was used. The weighted Hankel singular values are (1.5602, 0.4685, 0.0826, 0.0574, 0.0193, 0.0131, 0.0068, 0.0059) for N=3 and (1.5592, 0.4684, 0.0827, 0.0575, 0.0193, 0.0131, 0.0069, 0.0059) for N=10. The controller order was also reduced for the discretized system, which can be considered as fast-sampled and lifted system with N=1. The corresponding weighted Hankel singular values are (1.5539, 0.4660, 0.0817, 0.0568, 0.0191, 0.0130, 0.0068, 0.0059).

Fig. 4.5 shows the errors of approximation of the full-order continuous closed-loop transfer function by the transfer functions of the closed loops with reduced order controllers of order 2, obtained in the three above mentioned ways. The error is $|20 \log_{10}(\mathcal{T}/\mathcal{T}_r)|$ where $\mathcal{T}$ and $\mathcal{T}_r$ denote the closed-loop transfer functions with full order and with reduced order controllers. Clearly, with this measure, it is desirable (though not in general possible) for the error to be zero.

The results show that when fast-sampling is used during the reduction process, a superior result is obtained. It is sufficient in this case to use N=3 as the fast sampling rate. This corresponds to an angular frequency of approximately 190 rad/sec; the improvement in matching of $\mathcal{T}$ and $\mathcal{T}_r$ is evident starting at about 10 rad/sec.



1 reduction using simple discretization (no fast sampling)

2 reduction using fast-sampling and lifting for N=3

3 reduction using fast-sampling and lifting for N=10

Figure 4.5: Errors of approximation of the full-order closed loop by the closed loops with the reduced 2-nd order controllers.

Needless to say, whatever the sampling frequency is, it makes sense, especially if there is a problem with stability of the sampled-data closed loop, to use a more sophisticated scheme for obtaining the original (high order) discrete controller transfer function. (Keller and Anderson, 1991, 1992)

4.5 Conclusions

The proposed method allows one to reduce a discrete-time controller which is used in a closed loop with a continuous-time plant, sampler, zero-order hold and antialiasing filter. This reduction is based on information describing the system's behaviour not only at the sampling instants, but in inter-sample periods as well, and aims to preserve the closed-loop behaviour of the sampled-data loop. To get information about the inter-sample behaviour of the system, fast-sampling has been applied, followed by a lifting operation, which gives a time-invariant system. Obviously, the fast sampling procedure incurs an approximation error.

In the whole reduction process, there are actually three different types of error: (i) The error due to replacing a hybrid system by a multi-rate sampled-data system – this error can be made as small as desired by choosing a fast enough sampling rate for the faster of the two rates (ii) the error involved in replacing the problem of matching closed-loop transfer functions by the problem of matching (with weights) the open-loop responses of the controller – this error arises from neglecting second order terms, and has the potential to lead to a (mildly) less than optimal result in terms of closed-loop matching (iii) the error associated with approximating a high order transfer function by a low order one – this is obviously unavoidable.

The feasibility, efficiency and advantage of the proposed method have been confirmed by a practical numerical example.

The aim of science is not to open the door to infinite wisdom, but to set a limit to infinite error.

Bertolt Brecht *Life of Galileo*

Chapter 5

Error Bound for Transfer Function Order Reduction Using Frequency-Weighted Balanced Truncation

5.1 Introduction

Lower and, more importantly, upper frequency domain error bounds for the balanced truncation approximation in the non-weighted continuous-time case are well known and have been described in Enns (1984 a,b) and Glover (1984). However, no error bound formula has been available for the balanced truncation frequency-weighted problem.

The contribution of this Chapter is that an upper error bound for frequency weighted balanced controller reduction is obtained.[1] The bound is valid for both one-sided (input or output) and two-sided weighted balancing approxima-

1. Of course, it requires the reduced order system (corresponding to a stable high order system) to also be stable.

tion for stable weights which can otherwise be arbitrary.

An outline of this Chapter is as follows. In Section 5.2 the algorithm for frequency-weighted balanced reduction in continuous time is reviewed. The main result, the error bound formula itself, is presented in Section 5.3. An example (showing tightness of the bound) is given in Section 5.4, followed by some concluding remarks in Section 5.5.

5.2 Background

In this section the algorithm for non-weighted balanced reduction and the error bound are reviewed. Also, the algorithm for balanced weighted reduction is recalled.

Let us consider a stable transfer function K, given by a minimal state-space realization:

$$K(s) = C\,(sI - A)^{-1}\,B + D \tag{5.2.1}$$

Also, consider a stable input weight V(s) and a stable output weight W(s), realised in their minimal state-space form as:

$$V(s) = C_V\,(sI - A_V)^{-1}\,B_V + D_V \tag{5.2.2}$$

$$W(s) = C_W\,(sI - A_W)^{-1}\,B_W + D_W. \tag{5.2.3}$$

Then the weighted reduction problem is to find a stable lower-order transfer function K_r (of order r), such that the norm $\|W(s)\,[K(s) - K_r(s)]\,V(s)\|_\infty$ is minimal, or at least is approximately minimal.

A key application of interest is when K(s) is a controller and V(s) and W(s) are obtained by one of the methods described in Anderson and Liu (1989). (In a number of the methods of Anderson and Liu, 1989, V(s) or W(s) is the identity). Such a controller may be open loop unstable; the scheme presented here is restricted to reducing the order of the stable part of K(s), to yield the stable part of $K_r(s)$; the unstable part of K(s) is copied with $K_r(s)$.

Let us recall first the non-weighted case.

Definition 5.1. Given an n-th order, linear time invariant, asymptotically stable system with transfer function matrix K(s), a minimal realization of $K(s)=C(sI-A)^{-1}B+D$ is internally balanced if {A,B,C} satisfy the following Lyapunov equations:

$$\Lambda A^T + A \Lambda + B B^T = 0 \tag{5.2.4a}$$

$$\Lambda A + A^T \Lambda + C^T C = 0 \tag{5.2.4b}$$

and

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n), \tag{5.2.5}$$

where $\lambda_i \geq \lambda_{i+1} > 0, \qquad i=1,2, \ldots, n-1.$

In equation (5.2.4a), Λ is the controllability gramian, and in (5.2.4b), Λ is the observability gramian. Thus, a system is balanced when its controllability and observability gramians are equal and have a diagonal form.

Partition the system {A,B,C} and Λ as

$$A=\begin{pmatrix} A_r & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad B=\begin{pmatrix} B_r \\ B_2 \end{pmatrix} \qquad C=\begin{pmatrix} C_r & C_2 \end{pmatrix} \qquad \Lambda=\begin{pmatrix} \Lambda_r & 0 \\ 0 & \Lambda_2 \end{pmatrix} \tag{5.2.6}$$

where A_r, $\Lambda_r \in \mathbb{R}_{r\times r}$, $B_r \in \mathbb{R}_{r\times p}$, $C_r \in \mathbb{R}_{m\times r}$ and $r < n$.

Then the reduced-order system {A_r, B_r, C_r} is a good approximation of the system {A, B, C} if $\lambda_r >> \lambda_{r+1}$. In fact, the following two properties are true:

Lemma 5.1. (Pernebo & Silverman, 1982)
For a balanced asymptotically stable system {A, B, C} satisfying (5.2.1), and with Λ in the form of (5.2.5) satisfying (5.2.4) and partitioned as in (5.2.6), if $\lambda_r > \lambda_{r+1}$, then both subsystems {$A_r$, B_r, C_r} and {A_{22}, B_2, C_2} are asymptotically stable.

Lemma 5.2. (Enns, 1984b, Glover, 1984)

With the same hypothesis as Lemma 5.1, there holds a frequency error bound:

$$\| C\,(j\omega I-A)^{-1} B - C_r\,(j\omega I-A_r)^{-1} B_r \|_\infty \le 2(\lambda_{r+1} + \ldots + \lambda_n) = 2\,\mathrm{tr}(\Lambda_2) \tag{5.2.7}$$

Now, consider asymptotically stable frequency-weighting functions and associated minimal state-variable realizations $W(s)=C_w(sI-A_w)^{-1}B_w + D_w$ and $V(s)=C_v(sI-A_v)^{-1}B_v + D_v$. The basic idea is to change the gramians to reflect the introduction of the frequency weighting, to diagonalize these "weighted" gramians and then to truncate.

The frequency-weighted transfer function W(s)K(s)V(s) has a representation with the following state-space matrices:

$$\overline{A} = \begin{pmatrix} A_W & B_W C & B_W D C_V \\ 0 & A & B C_V \\ 0 & 0 & A_V \end{pmatrix} \quad \overline{B} = \begin{pmatrix} B_W D D_V \\ B D_V \\ B_V \end{pmatrix}$$

$$\overline{C} = (C_W \quad D_W C \quad D_W D C_V) \qquad \overline{D} = D_W D D_V$$

Let $\overline{P} = \begin{pmatrix} P_W & P_{12} & P_{13} \\ P_{12}^T & P & P_{23} \\ P_{13}^T & P_{23}^T & P_V \end{pmatrix}$ and $\overline{Q} = \begin{pmatrix} Q_W & Q_{12} & Q_{13} \\ Q_{12}^T & Q & Q_{23} \\ Q_{13}^T & Q_{23}^T & Q_V \end{pmatrix}$

be the solutions of the following Lyapunov equations:

$$\overline{P}\,\overline{A}^T + \overline{A}\,\overline{P} + \overline{B}\,\overline{B}^T = 0 \tag{5.2.8a}$$

$$\overline{Q}\,\overline{A} + \overline{A}^T\,\overline{Q} + \overline{C}^T\,\overline{C} = 0 \tag{5.2.8b}$$

Now, P and Q can be regarded as the frequency-weighted controllability and observability gramians for the original transfer function K(s). For later reference, we note that P_V is determined for A_V, B_V alone and Q_W is determined for A_W, C_W alone.

Consider a coordinate basis change to {A, B, C} which makes $P_{new} = Q_{new} = \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, $\sigma_i \ge \sigma_{i+1}$, $i=1,2,\ldots,n-1$. This new realization {A, B, C} is called a frequency-weighted balanced realization. (The coordinate basis change is easy to determine).

Partition {A, B, C} as in (5.2.6) and Σ as

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & \Sigma_2 \end{pmatrix} \tag{5.2.9}$$

where $\Sigma_r \in \mathbb{R}_{r\times r}$ and $r < n$.

Now, as previously, the reduction is achieved by eliminating the rows and columns of A, B and C corresponding to Σ_2 in Σ. The reduced order transfer function of order r is given by

$$K_r = C_r(sI - A_r)^{-1}B_r + D \tag{5.2.10}$$

A detailed, computer oriented description of this weighted balanced truncation algorithm is given in Integrated Systems Inc. (1991).

5.3 Main Result

The major aim of this section is to derive an upper error bound for balanced frequency-weighted controller reduction. This result is stated in the following theorem:

Theorem 5.1.
Let K(s), V(s) and W(s) be a stable transfer function of order n and stable weighting functions, respectively. The minimal state-space realizations are given by (5.2.1), (5.2.2) and (5.2.3), respectively. Also, let $K_r(s)$ (given by (5.2.10)) be a reduced order transfer function of order r, obtained by the frequency-weighted (with the weights V(s) and W(s)) balanced reduction technique. Assume that K_r is stable, which is guaranteed if V(s) or W(s) is constant, see Enns (1984 a,b). Then the following error bound holds [compare with (5.2.7)]:

$$\|W(s)\ [K(s) - K_r(s)]\ V(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \sqrt{\sigma_k^2 + (\alpha_k+\beta_k)\sigma_k^{3/2} + \alpha_k\beta_k\sigma_k}\ ,$$

where

$$\alpha_k = \|\Xi_{k-1}\|_\infty \; \|C_V \, \Phi_V \, P_V^{1/2}\|_\infty$$

and

$$\beta_k = \|Q_W^{1/2} \, \Phi_W \, B_W\|_\infty \; \|\Gamma_{k-1}\|_\infty$$

and

$$\Xi_{k-1}(s) = A_{21}{}^{k-1}\phi_{k-1}(s)B_{k-1} + b_k,$$

$$\Gamma_{k-1}(s) = C_{k-1}\phi_{k-1}(s)A_{12}{}^{k-1} + c_k,$$

$$\phi_{k-1}(s) = (sI-A_{k-1})^{-1}, \qquad \Phi_W(s) = (sI - A_W)^{-1}, \qquad \Phi_V(s) = (sI - A_V)^{-1},$$

$$A_k = \begin{pmatrix} A_{k-1} & A_{12}{}^{k-1} \\ A_{21}{}^{k-1} & a_{kk} \end{pmatrix}, \qquad B_k = \begin{pmatrix} B_{k-1} \\ b_k \end{pmatrix}, \qquad C_k = (\, C_{k-1} \;\; c_k)$$

and b_k and c_k are the k–th row of B_k and the k–th column of C_k, respectively and $A_n = A$, $B_n = B$, $C_n = C$.

For proof, see the Appendix H.

Remark 5.1. The assumption of stable K_r in the statement of the theorem cannot be omitted. It has long been thought that K_r is stable in the case of double–sided frequency–weighted balanced truncation (although proof existed for single–sided weighting only). Later in the thesis we will show that in the case of double–sided weighted balanced truncation the reduced controller K_r may be unstable.

Remark 5.2. Since the $L\infty$ bound on the error is expressed in terms of other $L\infty$ bounds, it might be thought that the advantage of the bound is minor. Several points should however be noted:

- The order of the transfer functions Ξ_{k-1} and Γ_{k-1} (viz. k–1 for $k = r + 1, \ldots, n$) will often be much less than that of $W(s)[K(s) - K_r(s)]V(s)$, viz. $(n + r)$ + deg W + deg V. Accordingly, the L_∞ bounds will be much easier to compute.

- The transfer functions $C_V \, \Phi_V \, P_V^{1/2}$ and $Q_W^{1/2} \, \Phi_W \, B_W$ are independent of K(s), depending just on the weights V(s) and W(s), and so their norms only need to be computed once.

- In the light of the above points, the bound formula lends itself to easy examination of a number of different trial values for r, leading to a subsequent selection.

Remark 5.3. The parameters α_k and β_k are finite. Indeed, $\|\Xi_{k-1}\|_\infty$ and $\|\Gamma_{k-1}\|_\infty$ are finite since the reduced controller is stable. Also, $\|\Phi_V\|_\infty$ is bounded since the weight V is stable and the unique solution P_V of the 3–3 block of Lyapunov equation (5.2.8a) is bounded since (A_V, B_V) is a controllable pair. Therefore, $\|C_V\ \Phi_V\ P_V^{1/2}\|_\infty$ is bounded. Furthermore, $\|C_V\ \Phi_V\ P_V^{1/2}\|_\infty$ depends on the input weight V only since P_V depends on A_V and B_V only. Similarly, $\|Q_W^{1/2}\ \Phi_W\ B_W\|_\infty$ is bounded and depends on the output weight W only.

It is easy to check also that the quantities $C_V\ \Phi_V\ P_V^{1/2}$ and $Q_W^{1/2}\ \Phi_W\ B_W$ do not depend on the coordinate basis choice for V(s) and W(s).

We can actually express a bound on $\|C_V\ \Phi_V\ P_V^{1/2}\|_\infty$ by using the upper bound of the solution P_V of the Lyapunov equation. As has been shown in Mori *et al.* (1986) and Troch (1987), $P_V \leq \|G\|_2\ S\ S^T$, where G is a positive definite symmetric matrix dependent on A_V only, and $S=[B_V\ \ A_VB_V\ \ \ldots\ A_V^{N_V-1}B_V]$ is the controllability matrix of $\{A_V, B_V\}$. Thus, the norm $\|C_V\ \Phi_V\ P_V^{1/2}\|_\infty$ can be bounded in terms of A_V, B_V and C_V as follows:

$$\|C_V\ \Phi_V\ P_V^{1/2}\|_\infty \leq \|G\|_2^{1/2}\ \|C_V\ \Phi_V\ S\|_\infty.$$

Similar result can be derived for a bound of the norm $\|Q_W^{1/2}\ \Phi_W\ B_W\|_\infty$ in terms of A_W, B_W and C_W.

We can also argue that if a different frequency weighted balanced realization is used, the same norm $\|\Xi_{k-1}\|_\infty$ results. Indeed, the frequency-weighted balanced realization is unique to within a sign change of a state variable when the singular values of the balanced gramian are distinct. Different balanced realizations are related by a transformation $T_k = \mathrm{diag}[t_i, i = 1, 2, \ldots, k]$ and $t_i = \pm 1$. It is not hard to conclude from this that $\|\Xi_{k-1}\|_\infty$ is invariant under a transformation T_k.

Actually, examples suggest that $\|\Xi_{n-1}\|_\infty$ and $\|\Gamma_{n-1}\|_\infty$ are bounded by quantities proportional to $\sqrt{\sigma_n}$ as $\sigma_n \to 0$. However, a proof of this has yet to be established. If true, the bound formula of the Theorem would depend on σ_k linearly, for $k = r+1, \ldots, n$.

Many frequency-weighted approximation problems have just a one-sided weighting. Then the main result becomes:

Corollary 5.1.
Let K(s), V(s) and W(s) be a stable controller of order n and stable weighting functions. Then a reduced order transfer function $K_r(s)$ of order r, obtained by single-sided frequency weighting (with either input weight V(s) or output weight W(s)) is stable (see Enns, 1984a) and the following error bounds are true:

$$\|[K(s) - K_r(s)]\, V(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \sqrt{\sigma_k^2 + \alpha_k \sigma_k^{3/2}}\ ,$$

$$\|W(s)\, [K(s) - K_r(s)]\|_\infty \leq 2 \sum_{k=r+1}^{n} \sqrt{\sigma_k^2 + \beta_k \sigma_k^{3/2}}\ .$$

Remark 5.4. In the non-weighted case, when W(s) = V(s) = I,

$$\|K(s) - K_r(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \sigma_k.$$

5.4 Examples

We now present three examples to illustrate how the bound on the weighted controller reduction error, obtained in accordance with the above theorem, compares to the actual weighted controller reduction error.

5.4.1 Example 5.1

This example has been studied in Enns (1984b). The stable controller to be reduced is given by its transfer function:

$$K(s) = (s^2 + 2.8s + 1.6)/(s^3 + 2.9s^2 + 3.1s + 1.5).$$

The stable input weighting is in the form:

$$V(s) = (s^3 + 2.9s^2 + 3.1s + 1.5)/(s^3 + 3.8s^2 + 4.4s + 1.6)$$

and there is no output weighting.

The weighted Hankel singular values are (0.53999, 0.12355, 0.0042758) and the reduced controllers of order 2 and 1 are

$$K_2(s) = 1.0135(s + 1.1373)/(s^2 + 1.3384s + 1.0715)$$

and

$$K_1(s) = 1.1694/(s + 0.83068)$$

respectively.

The actual weighted controller reduction errors are:

$$E_2 = \|[K(s) - K_2(s)]V(s)\|_\infty = 0.0085342$$

and

$$E_1 = \|[K(s) - K_1(s)]V(s)\|_\infty = 0.31977$$

for the reduced controllers of order 2 and 1 respectively.

When we estimate these values by calculating upper bounds of the errors using Theorem 5.1, we have:

$$\|C_V\ \Phi_V(j\omega)\ P_V^{1/2}\|_\infty = 0.31911,$$

$$\|\Xi_2(j\omega)\|_\infty = 0.18476,$$

$$\|\Xi_1(j\omega)\|_\infty = 0.75851,$$

$$\alpha_2 = 0.24205,$$

$$\alpha_3 = 0.058959$$

and the bounds are

$$\overline{E}_2 = 2(\sigma_3^2 + \alpha_3\ \sigma_3^{3/2})^{1/2} = 0.011793,$$

$$\bar{E}_1 = \bar{E}_2 + 2(\sigma_2^2 + \alpha_2\, \sigma_2^{3/2})^{1/2} = 0.33290.$$

Comparing the actual error values with their calculated upper bounds, we can see that the bound differs 4% from its actual value for the reduced controller of order 1, and the difference is 38% for the reduced controller of order 2. Thus, we can say that the approximated values are close to the real ones.

5.4.2 *Example 5.2*

As a second example, we consider a plant, given by its transfer function:

$$G(s) = (s + 0.8)(s + 2)/(s + 1.5)(s^2 + 1.4s + 1)$$

or, in a state-space form:

$$A_g = \begin{pmatrix} -2.9 & -3.1 & -1.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, B_g = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, C_g = (1 \quad 2.8 \quad 1.6), D_g = 0.$$

An LQG compensator was designed to control this plant. There were used state weighting $Q_C = I_3 + 100C_g^TC_g$, control weighting $R_C = 1$, state noise covariance $Q_f = I_3 + 100B_gB_g^T$ and measurement noise covariance R_f=1.

The design procedure resulted in the controller

$$K(s) = 10.3544(s + 1.86183)(s + 0.745649)$$
$$/(s + 19.8229)(s + 2.00134)(s + 0.800627).$$

Weighting functions, chosen to preserve (as far as possible) the closed-loop transfer function (see e.g. Anderson and Liu, 1989, Integrated Systems Inc., 1991) are in the form:

input weighting

$$V(s) = (s + 0.80062709)(s + 1.5)(s + 2.00134)(s + 19.8229)(s + 1.4s + 1)$$
$$\div[(s + 0.800687)(s + 1.30002)(s + 2.00147)(s + 19.279)$$
$$\times(s^2 + 2.14368s + 1.75884)],$$

output weighting

$W(s) = (s + 19.8229)(s + 2.00134)(s + 2)(s + 0.800627)(s + 0.8)$
$\div[(s + 0.800687)(s + 1.30002)(s + 2.00147)(s + 19.279)$
$\times(s^2 + 2.14368s + 1.75884)]$.

(Because K(s) is scalar, we could work with a single-sided weighting V(s)W(s). However, our goal here is to indicate the effect of two-sided weighting).

The weighted Hankel singular values σ are (0.052428, 0.011097, 0.00048095) and the weighted balanced reduced controllers of order 1 and 2 are

$$K_1(s) = 10.372/(s + 21.312)$$

and

$$K_2(s) = (10.384s + 11.916)/(s^2 + 21.299s + 26.205)$$

respectively.

The poles of the transfer functions with the controllers of reduced order are

$$(-19.4486, -1.0836 \pm 0.7547j, -1.2917 \pm 0.2295j)$$

and

$$(-20.8123, -1.29, -1.0546 \pm 0.8346j)$$

for the loops with the 2nd order and 1st order controllers respectively. Thus, the closed loops are stable.

The actual weighted balanced controller reduction errors are:

$$E_1 = \|W(s)[K(s) - K_1(s)]V(s)\|_\infty = 0.016581$$

and

$$E_2 = \|W(s)[K(s) - K_2(s)]V(s)\|_\infty = 0.0010472$$

for the reduced controllers of order 1 and 2 respectively.

When we estimate these values by calculating upper bounds of the errors using Theorem 5.1, we have

$$\|C_V \Phi_V(j\omega) P_V^{1/2}\|_\infty = 0.22893,$$

$$\|Q_W^{1/2} \Phi_W(j\omega) B_W\|_\infty = 0.0023564,$$

$$\|\Xi_1(j\omega)\|_\infty = 0.20246,$$

$$\|\Xi_2(j\omega)\|_\infty = 0.054046,$$

$$\|\Gamma_1(j\omega)\|_\infty = 2.8548,$$

$$\|\Gamma_2(j\omega)\|_\infty = 0.8166,$$

$\alpha_2 = 0.046348,\ \ \alpha_3 = 0.012373,\ \ \beta_2 = 0.006727,\ \ \beta_3 = 0.0019243$

and the bounds are

$$\overline{E}_2 = 2(\sigma_3{}^2 + (\alpha_3 + \beta_3)\ \sigma_3{}^{3/2} + \alpha_3\ \beta_3\ \sigma_3)^{1/2} = 0.0012547,$$

$$\overline{E}_1 = \overline{E}_2 + 2(\sigma_2{}^2 + (\alpha_2 + \beta_2)\ \sigma_2{}^{3/2} + \alpha_2\ \beta_2\ \sigma_2)^{1/2} = 0.029353.$$

Comparing the actual error values with their calculated upper bounds, we can see that the bound differs 77% from its actual value for the reduced controller of order 1, and the difference is 80% for the reduced controller of order 2. Thus, we can conclude that the bound in the double-side weighted case is not as tight as for single-side weighting (Example 5.1). But, anyway, the approximation is very good.

5.4.3 Example 5.3

The third example deals with a controller having its poles close to the jω-axis.

Let us consider the stable controller to be reduced, given by its transfer function:

$$K(s) = (s^2 + 2.8s + 1.6) \\ /\ (s^5 + 2.911s^4 + 3.1319s^3 + 1.5341s^2 + 0.01653s + 0.000015).$$

The poles are $\{-1.5, -0.7 \pm j\, 0.71414, -0.01, -0.001\}$. Thus two poles approach the jω-axis very closely.

The stable input weighting is given in the form

$$V(s) = K^{-1}(s)\ /\ [(s + 1)^2(s + 2)].$$

The weighted Hankel singular values σ are
{797.19, 1.6265, 0.07408, 0.0004583} and the weighted balanced reduced controllers of order 1, 2, 3 and 4 are

$$K_1(s) = -0.151 \,/\, (s + 2.21\cdot 10^{-9}),$$

$$K_2(s) = (-0.154s + 0.876) \,/\, (s^2 + 0.00742s + 9.98\cdot 10^{-6}),$$

$$K_3(s) = (0.00985s^2 - 0.102s + 1.47) \,/\, (s^3 + 1.21s^2 + 0.0171s + 1.21\cdot 10^{-5}),$$

$$K_4(s) = (0.00145s^3 - 0.0138s^2 + 1.06s + 1.04) \,/\, (s^4 + 1.33s^3 + 0.992s^2 + 0.0108s + 9.78\cdot 10^{-6})$$

respectively.

The actual weighted balanced controller reduction errors
$E_i = \|[K(j\omega) - K_i(j\omega)]\, V(j\omega)\|_\infty$, $i = 1, 2, 3, 4$ are:

$$E_1 = 321.03,\quad E_2 = 0.13124,\quad E_3 = 0.06691,\quad E_4 = 0.0009187.$$

When we estimate these values by calculating upper bounds of the errors using results of Theorem 5.1, we have:

$$\|C_V\Phi_V(j\omega)P_V^{1/2}\|_\infty = 0.41013,$$

$$\|\Xi_1(j\omega)\|_\infty = 2.5504\cdot 10^9,$$

$$\|\Xi_2(j\omega)\|_\infty = 9.4683\cdot 10^4,$$

$$\|\Xi_3(j\omega)\|_\infty = 4.8518\cdot 10^4,$$

$$\|\Xi_4(j\omega)\|_\infty = 8.0189\cdot 10^3,$$

$\alpha_2 = 1.046\cdot 10^9$, $\alpha_3 = 3.8832\cdot 10^4$, $\alpha_4 = 1.9899\cdot 10^4$, $\alpha_5 = 3.2888\cdot 10^3$

and the bounds are

$$\overline{E}_4 = 2(\sigma_5^2 + \alpha_5\,\sigma_5^{3/2})^{1/2} = 0.35927,$$

$$\overline{E}_3 = \overline{E}_4 + 2(\sigma_4^2 + \alpha_4\ \sigma_4^{3/2})^{1/2} = 22.203,$$

$$\overline{E}_2 = \overline{E}_3 + 2(\sigma_3^2 + \alpha_3\ \sigma_3^{3/2})^{1/2} = 78.166,$$

$$\overline{E}_1 = \overline{E}_2 + 2(\sigma_2^2 + \alpha_2\ \sigma_2^{3/2})^{1/2} = 93,239.$$

As we can see, actual errors may be large when controller has poles close to the imaginary axis. Also in this case the upper bound, derived in Theorem 5.1 and based on the inequality

$$||\Xi_i(j\omega)\ C_V\ \Phi_V(j\omega)\ P_{23\ [i+1]}{}^T||_\infty \leq ||\Xi_i(j\omega)||_\infty\ ||C_V\ \Phi_V(j\omega)\ P_V{}^{1/2}||_\infty \sqrt{\sigma_{i+1}},$$

becomes very conservative.

Let us try to make the bound tighter by using the inequality

$$||\Xi_i(j\omega)\ C_V\ \Phi_V(j\omega)\ P_{23,\ [i+1]}{}^T||_\infty \leq ||\Xi_i(j\omega)||_\infty\ ||C_V\ \Phi_V(j\omega)\ P_{23,\ [i+1]}{}^T||_\infty\ .$$

That gives us the following reduction error bounds:

$\overline{E}_4{}^a = 0.015098$, $\overline{E}_3{}^a = 3.3371$, $\overline{E}_2{}^a = 18.906$, $\overline{E}_1{}^a = 51,388$.

As we can see, this approach gives us better bounds, though still very conservative.

The bounds are worse when $|\Xi_i(j\omega)|$ and $|C_V\ \Phi_V(j\omega)\ P_{23,\ [i+1]}{}^T|$ assume their maximum values at mutually distant frequencies.

Finally, let us try to improve the results by bounding the reduction errors by the norms

$$||\Xi_i(j\omega) C_V \Phi_V(j\omega) P_{23,\ [i+1]}{}^T||_\infty .$$

That gives us extremely tight reduction error bounds:

$\overline{E}_4{}^b = 0.00091892$, $\overline{E}_3{}^b = 0.074513$, $\overline{E}_2{}^b = 0.25299$, $\overline{E}_1{}^b = 321.25$.

In this case the bounds differ from the actual errors by 0.07%, 93%, 11% and 0.02% for reduced controllers of order 1, 2, 3 and 4 respectively.

5.5 Conclusions

The controller reduction approach discussed in this Chapter involves a frequency weighted error between the full and reduced order transfer functions. The weighted balanced reduction technique was applied to reduce the order of the transfer function. An error bound formula for frequency weighted balanced truncation was obtained.

Three examples were used to demonstrate the effectiveness of the error bound formula. The first example involves an input weighting, the second - double-sided weighting. The calculated error bounds were close to the actual error values for both examples.

The third example studies the case of a transfer function having its poles close to the imaginary axis. It has been shown that the bound derived in the main theorem of the Chapter is conservative, but a way to improve the bound has been shown as well. It is widely held that in the unweighted case, the standard bounds are also conservative when poles are close to the $j\omega$–axis.

Issues of interest which remain open are the possibility of improving the error bound formula and of determining cases when the bound is tight/weak. In Enns (1984b), the question is considered for the unweighted reduction problem of when the bound is tight and weak. It is probably weak when the weighted system's transfer function has alternating poles and zeros almost along the $j\omega$–axis and tight when there are alternating poles and zeros along the negative real axis. This conclusion may carry over to the weighted case. The examples allow no real definitive conclusion on this point, although they tend to support the carry–over hypothesis.

Aristotle maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify this statement by examining his wives' mouths.

Bertrand Russell *Impact of Science on Society*

Chapter 6

New Results on Frequency-Weighted Balanced Truncation

6.1 Introduction

In this Chapter, we show by examples that Enns' frequency-weighted balanced truncation technique (Enns, 1984a) may give an unstable reduced order model or may not give any reduced order model of a particular order. We then propose a new frequency weighted balanced truncation technique which is guaranteed to yield stable reduced order models even when both input and output weightings are included. The proposed technique is essentially a simple generalisation of Lin and Chiu's technique (Lin and Chiu, 1992) and can handle weighting transfer functions which are proper. Furthermore, we also present frequency response error bounds for the proposed technique.

Several examples are presented to compare the Enn's technique and the new scheme. Sometimes, the Enns' scheme gives a superior error performance and sometimes the proposed technique gives a superior performance.

6.2 Some Remarks on Enns' Technique

In this section, we show by different examples that the frequency weighted balanced truncation technique (Enns, 1984a) when both input and output weightings are included (i) may yield unstable reduced order models or (ii) may not yield any reduced-order model of a particular order.

6.2.1 Example A

Consider the third-order system

$$K(s)=(8s^2+6s+2)/(s^3+4s^2+5s+2)$$

with the poles and zeros at $p_i=-1, -1, -2$ and $z_i=-0.3750\pm j\,0.3307$ respectively.

The input and output weights are respectively:

$$V(s)=1/(s+3) \quad \text{and} \quad W(s)=1/(s+4).$$

The diagonalized weighted controllability and observability Gramians are:

$$P=Q=\mathrm{diag}\{0.0513,\ 0.0417,\ 0.0057\}.$$

Enns' (1984a) method gives the following 1st and 2nd order models:

$$K_1(s)=-0.1563/(s-0.1085)$$

$$K_2(s)=(7.705s+3.3214)/(s^2+3.4056s+3.9040).$$

Clearly, $K_1(s)$ is unstable and $K_2(s)$ is stable.

6.2.2 Example B

Consider the third-order original system of Example A with the following input and output weights respectively:

$$V(s)=1/(s+5.72624615) \text{ and } W(s)=1/(s+4).$$

The diagonalized weighted controllability and observability Gramians are:

$$P=Q=\mathrm{diag}\{0.0286,\ 0.0265,\ 0.0032\}.$$

Enns' (1984a) method gives the following 1st and 2nd order models:

$$K_1(s)=7.0102\times 10^{-9}/(s+3.8275\times 10^{-9})$$

$$K_2(s)=(7.7761s+3.2742)/(s^2+3.4506s+3.8724).$$

Note both $K_1(s)$ and $K_2(s)$ are stable. However, $K_1(s)\approx 0$, which suggests that a 1st order model does not exist!

6.2.3 Example C

Following is a discrete-time example which yields an unstable pole in the reduced order system. If the weights are slightly adjusted, the unstable pole can be moved onto the unit circle (to $z=-1$); in this case, the residue is not zero.

Consider the 4th order system of Wagie and Skelton (1986):

$$K(z)=z^3/(z^4+1.1z^3-0.01z^2-0.275z-0.06)$$

with the following weights

$$W(z)=V(z)=(z+0.9)/(z+0.1).$$

The diagonalized weighted controllability and observability Gramians are:

$$P=Q=\mathrm{diag}\{1.1439,\ 0.3106,\ 0.2391,\ 0.0032\}.$$

The first-order model obtained by Enns' (1984a) technique is

$$K_1(z)=1.0241/(z+1.0221),$$

which is clearly unstable.

6.3 Generalization and Error Bounds

In this section, we generalize Lin and Chiu's (1992) technique to handle proper weighting functions. We also derive frequency response error bounds for the generalized technique presented.

6.3.1 Generalized Frequency-Weighted Technique

Let the transfer function of the original stable system be given by

$$K(s)=C(sI-A)^{-1}B+D,$$

where $\{A,B,C,D\}$ is a minimal state-space realization. Let the transfer functions of the stable input and output weights be respectively

$$V(s)=C_V(sI-A_V)^{-1}B_V+D_V$$

$$W(s)=C_W(sI-A_W)^{-1}B_W+D_W$$

where $\{A_V,B_V,C_V,D_V\}$ and $\{A_W,B_W,C_W,D_W\}$ are minimal realizations. The state-space realization of the augmented system $K(s)V(s)$ is

$$\overline{A}_i=\begin{bmatrix} A & BC_V \\ 0 & A_V \end{bmatrix},\quad \overline{B}_i=\begin{bmatrix} BD_V \\ B_V \end{bmatrix},\quad \overline{C}_i=[C \quad 0].$$

The state-space realization of the augmented system $W(s)K(s)$ is

$$\overline{A}_o=\begin{bmatrix} A & 0 \\ B_WC & A_W \end{bmatrix},\quad \overline{B}_o=\begin{bmatrix} B \\ 0 \end{bmatrix},\quad \overline{C}_o=[D_WC \quad C_W].$$

Let

$$\overline{P}_i=\begin{bmatrix} P & P_{12} \\ P_{12}^T & P_V \end{bmatrix},\quad \text{and}\quad \overline{Q}_o=\begin{bmatrix} Q & Q_{12}^T \\ Q_{12} & Q_W \end{bmatrix} \tag{6.3.1}$$

be the solutions of the following Lyapunov equations:

$$\overline{A}_i\overline{P}_i + \overline{P}_i\overline{A}_i^T + \overline{B}_i\overline{B}_i^T = 0 \tag{6.3.2}$$

$$\overline{A}_o^T\overline{Q}_o + \overline{Q}_o\overline{A}_o + \overline{C}_o^T\overline{C}_o = 0 \tag{6.3.3}$$

Assuming that there are no pole-zero cancellations in K(s)V(s) and W(s)K(s), the Gramians, $\overline{P}_i$ and $\overline{Q}_o$ are positive definite.

Theorem 6.1: Consider the system $\{A,B,C,D\}$ with input weight, $\{A_V,B_V,C_V,D_V\}$ and output weight, $\{A_W,B_W,C_W,D_W\}$. If

$$X=BD_V-P_{12}P_V^{-1}B_V$$

$$Y=D_WC-C_WQ_W^{-1}Q_{12}$$

where P_{12}, P_V, Q_{12} and Q_W are given by eqn. (6.3.1), then the realization $\{A,X,Y\}$ is minimal.

Proof: Let $\overline{T}_i$ and $\overline{T}_o$ be transformations which block diagonalize the Gramians, $\overline{P}_i$ and $\overline{Q}_o$, and have the following structure:

$$\overline{T}_i = \begin{bmatrix} I & P_{12}P_V^{-1} \\ 0 & I \end{bmatrix}, \quad \text{and} \quad \overline{T}_o = \begin{bmatrix} I & 0 \\ -Q_W^{-1}Q_{12} & I \end{bmatrix}$$

Since $\{A_V,B_V\}$ is controllable and $\{C_W,A_W\}$ is observable, P_V^{-1} and D_W^{-1} in the above equations exist. The block-diagonalized Gramians now have the following structure:

$$D_i = \overline{T}_i^{-1}\overline{P}_i\overline{T}_i^{-T} = \begin{bmatrix} P-P_{12}P_V^{-1}P_{12}^T & 0 \\ 0 & P_V \end{bmatrix}$$

and

$$D_o = \overline{T}_o^T\overline{Q}_o\overline{T}_o = \begin{bmatrix} Q-Q_{12}^TQ_W^{-1}Q_{12} & 0 \\ 0 & Q_W \end{bmatrix}$$

The corresponding state-space realizations have the following structures:

$$A_i=\overline{T}_i^{-1}\overline{A}_i\overline{T}_i = \begin{bmatrix} A & X_{12} \\ 0 & A_V \end{bmatrix}, \quad B_i=\overline{T}_i^{-1}\overline{B}_i = \begin{bmatrix} X \\ B_V \end{bmatrix}, \quad C_i=\overline{C}_i\overline{T}_i = [C \quad CP_{12}P_V^{-1}].$$

where

$$X_{12}=AP_{12}P_V^{-1}+BC_V-P_{12}P_V^{-1}A_V$$

$$X=BD_V-P_{12}P_V^{-1}B_V$$

$$A_o=\overline{T}_o^{-1}\overline{A}_o\overline{T}_o=\begin{bmatrix} A & 0 \\ Y_{12} & A_W \end{bmatrix},\quad B_o=\overline{T}_o^{-1}\overline{B}_o=\begin{bmatrix} B \\ Q_W^{-1}Q_{12}B \end{bmatrix},\quad C_o=\overline{C}_o\overline{T}_o=[Y \quad C_W].$$

where

$$Y_{21}=Q_W^{-1}Q_{12}A+B_WC-A_WQ_W^{-1}Q_{12}$$

$$Y=D_WC-C_WQ_W^{-1}Q_{12}$$

Note that the transformations $\overline{T}_i$ and $\overline{T}_o$ do not change the diagonal blocks of the system matrices $\overline{A}_i$ and $\overline{A}_o$. The new realizations now satisfy the following Lyapunov equations:

$$A_iD_i+D_iA_i^T+B_iB_i^T=0$$

$$A_o^TD_o+D_oA_o+C_o^TC_o=0$$

Since D_i and D_o are positive definite and A_i and A_o are stable, $\{A_i,B_i\}$ is controllable and $\{C_o,A_o\}$ is observable.

Expanding the (1,1) blocks of the Lyapunov equations (6.3.2)–(6.3.3), we get

$$A(P-P_{12}P_V^{-1}P_{12}^T)+(P-P_{12}P_V^{-1}P_{12}^T)A^T+XX^T=0$$

$$A^T(Q-Q_{12}^TQ_W^{-1}Q_{12})+(Q-Q_{12}^TQ_W^{-1}Q_{12})A+Y^TY=0$$

Since $(P-P_{12}P_V^{-1}P_{12}^T)$ and $(Q-Q_{12}^TQ_W^{-1}Q_{12})$ are positive definite and A is stable, it follows immediately that $\{A,X\}$ is controllable and $\{Y,A\}$ is observable or the realization $\{A,X,Y\}$ is minimal.

Remark 6.1: The Gramian $(P-P_{12}P_V^{-1}P_{12}^T)$ has the following interpretation. Consider the optimization problem of minimizing the input energy $\int_{-T}^{0} u^T(t)u(t)dt$ to the system

$$\begin{bmatrix} \dot{x}(t) \\ \dot{x}_v(t) \end{bmatrix} = \bar{A}_i \begin{bmatrix} x(t) \\ x_v(t) \end{bmatrix} + \bar{B}_i u(t)$$

under the constraint that $x(-T)=0$, $x_v(-T)=0$ and $x(0)=x_0$, $x_v(0)=x_{v0}$. When $T \rightarrow \infty$, the minimum energy is

$$[x_0^T \quad x_{v0}^T]\, \bar{P}_i^{-1} \begin{bmatrix} x_0 \\ x_{v0} \end{bmatrix}$$

Hence, under the additional constraint $x_{v0}=0$, the minimum is $x_0^T(P-P_{12}P_V^{-1}P_{12}^T)^{-1}x_0$.

We can interpret $(Q-Q_{12}^TQ_W^{-1}Q_{12})$ via a dual statement. Consider the unforced system

$$\begin{bmatrix} \dot{x}(t) \\ \dot{x}_w(t) \end{bmatrix} = \bar{A}_o \begin{bmatrix} x(t) \\ x_w(t) \end{bmatrix}$$

$$z(t) = \bar{C}_o \begin{bmatrix} x(t) \\ x_w(t) \end{bmatrix} + v(t)$$

where $v(t)$ is unit variance white noise. The mean square error covariance in estimating the initial state $[x^T(0)\ ,\ x_w^T(0)]^T$ from $z(t)$, $0 \leq t < \infty$, is $\bar{Q}_o^{-1}$ and $(Q-Q_{12}^TQ_W^{-1}Q_{12})^{-1}$ is the error covariance in estimating $x(0)$.

6.3.2 Generalized Algorithm

The new frequency weighted balanced truncation algorithm is based on diagonalizing the weighted Gramians, $(P-P_{12}P_V^{-1}P_{12}^T)$ and $(Q-Q_{12}^TQ_W^{-1}Q_{12})$ instead of the weighted Gramians P and Q.

1. Given the stable minimal realizations, $\{A,B,C,D\}$, $\{A_V,B_V,C_V,D_V\}$ and $\{A_W,B_W,C_W,D_W\}$, compute X and Y.

2. Calculate the transformation, T which balances $\{A,X,Y\}$. In other words, $T \in R^{n \times n}$ is a nonsingular matrix, such that

$$T^{-1}(P-P_{12}P_V^{-1}P_{12}^T)T^{-T} = T^T(Q-Q_{12}^TQ_W^{-1}Q_{12})T = \Sigma$$

where

$$\Sigma = \mathrm{diag}\{\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_r,\ \sigma_{r+1},\ \ldots,\ \sigma_n\}$$

and $\sigma_i \geq \sigma_{i+1}$, i=1,2, . . . , n−1.

3. Compute the frequency weighted balanced realization

$$\overline{A}=T^{-1}AT, \quad \overline{B}=T^{-1}B, \quad \overline{C}=CT$$

4. Partition $\{\overline{A},\overline{B},\overline{C}\}$ as follows:

$$\overline{A} = \begin{bmatrix} A_r & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \overline{B} = \begin{bmatrix} B_r \\ B_2 \end{bmatrix}, \quad \overline{C} = [C_r \quad C_2].$$

where $A_r \in R^{r \times r}$, $B_r \in R^{r \times p}$, $C_r \in R^{m \times r}$, and $r<n$.

5. The reduced order model obtained is $K_r(s)=C_r(sI-A_r)^{-1}B_r+D$.

Remark 6.2: In the case of input weighting alone, the realization {A,X,C} is balanced in step 2 instead of {A,X,Y}. Similarly the realization {A,B,Y} is balanced when only output weighting is used. The rest of the steps remain unaltered.

Remark 6.3: When $D_W=0$, $D_V=0$, $Q_W=I$ and $P_V=I$, the proposed algorithm reduces to the algorithm of Lin and Chiu (1992).

Remark 6.4: It can be easily verified that the above algorithm also holds good for the model reduction of discrete systems.

Remark 6.5: We know that the magnitude of singular values play a vital role in the approximation obtained using balanced reduction technique. Therefore, it is important to know how the magnitude of singular values obtained using the proposed technique compare with the magnitude of singular values obtained via Enns' technique. The following lemma gives the relationship between the singular values obtained using the two techniques in case of double-sided or single-sided weightings.

Lemma 6.1: If P and $(P-P_{12}P_V^{-1}P_{12}^T)$ are the weighted controllability Gramians, Q and $(Q-Q_{12}^TQ_W^{-1}Q_{12})$ are the weighted observability Gramians, P_K is the controllability Gramian (without weighting) and Q_K is the observability Gramian (without weighting), then

$$\text{(i)} \quad \lambda_i[PQ] \geq \lambda_i\,[(P-P_{12}P_V^{-1}P_{12}^T)(Q-Q_{12}^TQ_W^{-1}Q_{12})]$$

(ii) $\lambda_i[PQ_K] \geq \lambda_i\ [(P-P_{12}P_V^{-1}P_{12}^T)Q_K]$

(iii) $\lambda_i[P_KQ] \geq \lambda_i\ [P_K(Q-Q_{12}^TQ_W^{-1}Q_{12})]$

Proof: (i)

$$\lambda_i[PQ]=\lambda_i[Q^{1/2}PQ^{1/2}]$$

$$\geq \lambda_i[Q^{1/2}(P-P_{12}P_V^{-1}P_{12}^T)Q^{1/2}]$$

$$= \lambda_i[(P-P_{12}P_V^{-1}P_{12}^T)^{1/2}Q(P-P_{12}P_V^{-1}P_{12}^T)^{1/2}]$$

$$\geq\lambda_i[(P-P_{12}P_V^{-1}P_{12}^T)^{1/2}(Q-Q_{12}^TQ_W^{-1}Q_{12})(P-P_{12}P_V^{-1}P_{12}^T)^{1/2}]$$

$$= \lambda_i\ [(P-P_{12}P_V^{-1}P_{12}^T)(Q-Q_{12}^TQ_W^{-1}Q_{12})].$$

Here $[\cdot]^{1/2}$ denotes the positive definite symmetric square root of a positive definite matrix. The other two inequalities follow similarly.

6.3.3 Error Bounds

The main aim of this section is to derive the error bounds for the proposed frequency weighted balanced reduction technique. The result is similar to the one established in Chapter 5. As with the result in Chapter 5, the L_∞ bound on the error is expressed in terms of other L_∞ norms. However, the order of the transfer functions involved in the other norms are less, and sometimes very much less, than the order of the weighted error transfer function. Therefore, the L_∞ norms will be easier to compute than the L_∞ norm of the weighted error transfer function. To derive the error bounds, we define the following notation:

$$\phi_{k-1}(s)=(sI-A_{k-1})^{-1},$$

$$\Phi_V(s)=(sI-A_V)^{-1},$$

$$\Phi_W(s)=(sI-A_W)^{-1},$$

$$\Phi(s)=(sI-A)^{-1},$$

$$A_k=\begin{bmatrix} A_{k-1} & a_{12}^k \\ a_{21}^k & a_{kk} \end{bmatrix},\quad B_k=\begin{bmatrix} B_{k-1} \\ b_k \end{bmatrix},\text{ and}\quad C_k=[C_{k-1}\quad c_k]$$

where b_k and c_k are the k-th row of B_k and k-th column of C_k respectively and A_n=A, B_n=B and C_n=C.

Theorem 6.2: Let K(s) be a proper, stable transfer function of order n and V(s) and W(s) be proper and stable weighting functions. If $K_r(s)$ is a proper, stable reduced-order model obtained using the proposed frequency weighted balanced reduction technique, then the following error bound holds:

$$E_a = \|W(s)[K(s)-K_r(s)]V(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \{(\sigma_k+\alpha_k+\lambda_k)(\sigma_k+\beta_k+\omega_k)\}^{1/2} = E_b \tag{6.3.4}$$

where

$$\alpha_k=\|\Xi_{k-1}\|_\infty \ \|C_V \ \Phi_V \ (P_{12}^k)^T \ e_k^T \ \|_\infty,$$

$$\beta_k=\|e_k \ (Q_{12}^k)^T \ \Phi_W \ B_W \ \|_\infty \ \|\Gamma_{k-1}\|_\infty,$$

$\lambda_k=\|\Lambda_k\|_\infty$, $\omega_k=\|\Omega_k\|_\infty$, P_{12}^k=first k rows of P_{12}, Q_{12}^k=first k columns of Q_{12},

$\Lambda_{k-1}(s)=a_{21}^k \ \phi_{k-1}(s) \ P_{k-1} + p_k$, $\quad \Omega_{k-1}(s)=Q_{k-1} \ \phi_{k-1}(s) \ a_{12}^k + q_k$,

$\Xi_{k-1}(s)=a_{21}^k \ \phi_{k-1}(s) \ B_{k-1} + b_k$, $\quad \Gamma_{k-1}(s)=C_{k-1} \ \phi_{k-1}(s) \ a_{12}^k + c_k$,

$[P_{k-1} \quad p_k]^T = P_{12}^k \ P_V^{-1} \ (P_{12}^k)^T \ e_k$, $\quad [Q_{k-1} \quad q_k] = e_k^T \ (Q_{12}^k)^T \ Q_W^{-1} \ Q_{12}^k$,

e_k is the k-th column of k-th order identity matrix.

For proof, see the Appendix I.

Remark 6.6: When the reduced order models $K_r(s)$ are obtained by single-sided frequency weighting the following error bounds hold:

$$\|[K(s)-K_r(s)]V(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \{(\sigma_k+\alpha_k+\lambda_k)\sigma_k\}^{1/2}$$

with input weight V(s), where σ_k^2 are the eigenvalues of $[(P-P_{12}P_V^{-1}P_{12}^T)Q_K]$.

$$\|W(s)[K(s)-K_r(s)]\|_\infty \leq 2 \sum_{k=r+1}^{n} \{\sigma_k(\sigma_k+\beta_k+\omega_k)\}^{1/2}$$

with output weight W(s), where σ_k^2 are the eigenvalues of $[P_K(Q-Q_{12}^TQ_W^{-1}Q_{12})]$.

Remark 6.7: If the reduced order model $K_r(s)$ is obtained without frequency weighting, then W(s)=V(s)=I. The following result of Enns (1984a) can be obtained easily:

$$\|K(s)-K_r(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \sigma_k$$

where the Gramians, $P_K=Q_K=\text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$.

Remark 6.8: If the reduced order model is obtained via the frequency weighted technique of Enns (1984a), then in the eqn. (6.3.4), $\lambda_k=0$, and $\omega_k=0$. This is because P and Q are diagonalized instead of $P-P_{12}P_V^{-1}P_{12}^T$ and $Q-Q_{12}^TQ_W^{-1}Q_{12}$. Hence, the error bounds as shown in Chapter 5 can be obtained from (6.3.4) as follows:

$$\hat{E}_a = \|W[K-K_r]V\|_\infty \leq 2 \sum_{k=r+1}^{n} \{(\hat{\sigma}_k+\hat{\alpha}_k)(\hat{\sigma}_k+\hat{\beta}_k)\}^{1/2} = \hat{E}_b$$

where the weighted Gramians, $P=Q=\text{diag}\{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_n\}$. Formulas for $\hat{\alpha}_k$ and $\hat{\beta}_k$ are similar to ones already defined in Theorem 6.2.

6.4 Examples

In this section three examples are used to illustrate the proposed technique and compare with other techniques (Enns, 1984a, Al-Saggaf and Franklin, 1988).

6.4.1 Example 6.1

Consider the fourth-order system

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 5 \\ 1/2 & -3/2 \\ 1 & -5 \\ -1/2 & 1/6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 4/15 & 1 & 0 & 1 \end{bmatrix}$$

of Lin and Chiu (1992) with the following input and output weights:

$$A_V = A_W = \begin{bmatrix} -4.5 & 0 \\ 0 & -4.5 \end{bmatrix}, \quad C_V = C_W = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix},$$

$$B_V = B_W = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \text{ and } D_V = D_W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The frequency weighted balanced realization obtained using the proposed technique (with double-sided weighting) is:

$$A = \begin{bmatrix} -0.8390 & -0.5245 & 0.2259 & 0.0105 \\ 0.5100 & -2.4370 & 0.3638 & 0.5466 \\ -0.2899 & 1.4993 & -2.3925 & 1.0804 \\ 0.0136 & -0.6032 & -0.4143 & -4.3315 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.1356 & 1.7054 \\ 0.4573 & -1.3487 \\ -0.1279 & 1.2508 \\ 0.4838 & -0.2041 \end{bmatrix}, \quad C = \begin{bmatrix} 1.6843 & 1.4422 & -0.7353 & 0.0373 \\ 0.2914 & 0.5880 & 0.1387 & -0.6008 \end{bmatrix}.$$

The diagonalized Gramians are:

$$P - P_{12} P_V^{-1} P_{12}^T = Q - Q_{12}^T Q_W^{-1} Q_{12}$$

$$= \text{diag}\{5.0106, 0.2062, 0.0566, 0.0042\} \qquad (6.4.1)$$

Table 6.1: The error bounds for the models.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	2.5744	3.7380	45.20
2	0.5607	1.1493	104.97
3	0.1645	0.2277	38.40

The frequency weighted balanced realization obtained using the Enns' technique (with double-sided weighting) is:

$$A = \begin{bmatrix} -0.5763 & -1.0319 & 0.1138 & 0.0250 \\ 0.9397 & -3.1861 & 0.2495 & 0.3937 \\ -0.2788 & 1.7566 & -1.4629 & 2.7755 \\ 0.0339 & -0.6090 & -0.6890 & -4.7747 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2139 & 1.4360 \\ 0.4610 & -1.3669 \\ -0.1099 & 0.4860 \\ 0.4102 & -0.0088 \end{bmatrix}, \quad C = \begin{bmatrix} 1.4301 & 1.4232 & -0.2221 & 0.0331 \\ 0.2389 & 0.3535 & 0.2803 & -0.4468 \end{bmatrix}$$

The diagonalized Gramians are:

$$P=Q=\mathrm{diag}\{7.1449, 0.7924, 0.1397, 0.0399\}. \tag{6.4.2}$$

Comparing the singular values in eqn. (6.4.1) and (6.4.2), it is easy to see that lemma 6.1 holds. The reduced order realizations of order 1, 2, and 3 are all stable, with error bounds shown in Table 6.2.

Table 6.2: The error bounds for the models.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	2.1291	2.7018	26.90
2	0.2660	0.4605	73.13
3	0.1131	0.1243	9.90

The frequency weighted balanced realization obtained using the proposed technique (with input weighting alone) is:

$$A = \begin{bmatrix} -0.7129 & -0.7955 & 0.1718 & 0.0153 \\ 0.7364 & -2.8747 & 0.3537 & 0.4795 \\ -0.2940 & 1.6558 & -1.8239 & 1.8358 \\ 0.0022 & -0.6663 & -0.6572 & -4.5885 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.1933 & 1.0435 \\ 0.4493 & -1.7344 \\ -0.3340 & 1.2189 \\ 0.6655 & -0.2278 \end{bmatrix}, \quad C = \begin{bmatrix} 2.1132 & 1.0499 & -0.3106 & 0.0242 \\ 0.3689 & 0.4821 & 0.3192 & -0.2724 \end{bmatrix}$$

The diagonalized Gramians are:

$$P - P_{12} P_V^{-1} P_{12}^T = Q_K$$

$$= \mathrm{diag}\{3.2275, 0.2322, 0.0544, 0.0081\}. \qquad (6.4.3)$$

Table 6.3: The error bounds for the models.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	1.1984	1.7232	43.79
2	0.2089	0.4386	109.89
3	0.0706	0.0785	11.20

The frequency weighted balanced realization obtained using the Enns' technique (with input weighting alone) is:

$$A = \begin{bmatrix} -0.5364 & -1.0800 & 0.0966 & 0.0259 \\ 1.0085 & -3.2648 & 0.2127 & 0.3517 \\ -0.2679 & 1.7418 & -1.4426 & 2.9429 \\ 0.0510 & -0.5614 & -0.6392 & -4.7562 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.1713 & 0.9929 \\ 0.3588 & -1.0874 \\ -0.1174 & 0.3586 \\ 0.3302 & -0.0057 \end{bmatrix}, \quad C = \begin{bmatrix} 1.9811 & 1.7296 & -0.2412 & 0.0352 \\ 0.3318 & 0.4350 & 0.4084 & -0.4997 \end{bmatrix}$$

The diagonalized Gramians are:

$$P = Q_K = \mathrm{diag}\{3.7613, 0.4871, 0.0780, 0.0264\}. \qquad (6.4.4)$$

Comparing the singular values in eqn. (6.4.3) and (6.4.4), it is easy to see that lemma 6.1 holds.

Table 6.4: The error bounds for the models.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	1.1310	1.4002	23.81
2	0.1342	0.2356	75.60
3	0.0654	0.0674	3.14

6.4.2 *Example 6.2*

Consider the original system of sub-sections 6.2.1 and 6.2.2 with the following input and output weights:

$$V(s)=1/(s+5.8) \quad \text{and} \quad W(s)=1/(s+4)$$

Enns' technique gives stable 1st and 2nd order models for this example. The following tables give a comparison of errors and error bounds for the models obtained by the proposed and Enns' techniques.

Table 6.5: The actual errors and error bounds for the models obtained by the proposed technique.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	0.0885	0.2257	155.14
2	0.0074	0.0087	18.35

Table 6.6: The actual errors and error bounds for the models obtained by Enns' technique.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	0.0883	0.3828	333.60
2	0.0067	0.0070	4.34

Comparison of Tables 6.5 and 6.6, show that although there are more terms in the error bound formula for the proposed technique than the Enns' technique, it is possible to get lower error bounds with the proposed technique than with the Enns' technique.

6.4.3 Example 6.3

In this section, for the sake of completeness we compare the proposed technique with Al-Saggaf and Franklin's (1988) technique. Note that Al-Saggaf and Franklin's technique is guaranteed to yield stable reduced order models. However, it can be applied to the case of single-sided weighting only, and furthermore, the technique makes the following assumptions: (i) the weighting system is strictly proper and (ii) C_V^{-1} exists (in case of input weighting) or B_W^{-1} exists (in case of output weighting). (For a scalar system, this means that the weights must be first order).

For comparison of the techniques, we consider the example of sub-section 6.4.1 with

$$D_V = D_W = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Table 6.7: The actual errors and error bounds for the models obtained (with input weight alone) by the proposed technique of this paper.

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	0.5605	0.7388	31.81
2	0.0619	0.1353	118.74
3	0.0314	0.0345	9.87

Table 6.8: The actual errors and error bounds for the models obtained (with input weight alone) by Enns' technique

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	0.5568	0.6973	25.23
2	0.0620	0.1182	90.55
3	0.0322	0.0336	4.40

Table 6.9: The actual errors and error bounds for the models obtained (with input weight alone) by Al-Saggaf and Franklin's technique

Order of the model	Actual Error E_a	Error Bound E_b	$\lvert(E_a-E_b)/E_a\rvert \times 100$
1	2.4672	3.1994	29.68
2	0.7387	0.8278	12.06
3	0.0662	0.0662	0.00

The above tables show that Al-Saggaf and Franklin's technique results in significantly poorer model approximations than arise with the proposed and the Enns' procedures, although the actual errors are closer to the error bounds, i.e. the error bounds are tighter.

6.5 Conclusions

It was observed during simulations that the results (actual errors and error bounds) obtained by the proposed technique were closer to the results obtained by Enns' technique when the poles of the weights were away from jω–axis compared to the poles of the original system. This is because when the poles of weights are very much away from the jω–axis compared to the poles of the original system, the elements of the matrices P_{12} and Q_{12} become very small. Therefore, the weighted Gramians of the two techniques will be approximately equal, i.e., $P-P_{12}P_V^{-1}P_{12}^T \approx P$, and $Q-Q_{12}Q_W^{-1}Q_{12}^T \approx Q$. The actual errors and the error bounds of the two techniques will also be approximately equal.

There is yet another reduction scheme which can be contemplated, intermediate in a sense between the Enns' scheme and the scheme of this paper. In this intermediate scheme, we balance $P-P_{12}P_V^{-1}P_{12}^T$ and Q (or $Q-Q_{12}Q_W^{-1}Q_{12}^T$ and P). The resulting reduced order system is guaranteed to have no poles in Re[s]>0; we have however constructed an example where the reduced order system has a pole at Re[s]=0.

You can judge your age by the amount of pain you feel when you come in contact with a new idea.

John Nuveen

Chapter 7

Multiplicative Approximation of Transfer Functions with Frequency Weighting

7.1 Introduction

In this Chapter we consider a class of frequency weighted model reduction which finds an r–th order transfer function $V_r(s)$ aimed at minimizing the following error index:

$$E_W = \|V^{-1}[V-Vr]W\|_\infty \qquad (7.1.1)$$

Here V(s) is a given stable transfer function of order greater than r, and W(s) is a given stable weighting function. With W(s) the identity, balanced stochastic truncation (BST) (Desai and Pal, 1984, Green 1988 a,b, Green and Anderson, 1990) is one method that can be used to find a $V_r(s)$ which is approximately minimizing. Our scheme with non–constant W(s) generalizes BST and we term it therefore weighted BST. We also term E_W of (7.1.1) the weighted multiplicative error.

The principal contribution of this Chapter is to develop an algorithm for the input-weighted balanced stochastic truncation method. This algorithm can be extended to the output-weighted and the two-sided weighted cases.

An outline of this Chapter is as follows. In Section 7.2, the algorithm for non-weighted balanced stochastic truncation is reviewed and the algorithm for the input-weighted case is presented. An example is given in Section 7.3, followed by some concluding remarks in Section 7.4.

7.2 Algorithm for Input-Weighted BST

In this section, the algorithm for non-weighted balanced stochastic truncation is reviewed (Desai and Pal, 1984, Green, 1988 a,b, Green and Anderson, 1990) and then the algorithm for the input-weighted case is presented.

Consider a possibly non-minimum phase, stable n-th order (n>r) square transfer function matrix V(s), given by a minimal state-space realization

$$V(s) = D_v + H(sI-A)^{-1}B \qquad (7.2.1)$$

with $\det[V(j\omega)]\neq 0$ for all $\omega\in[0,\infty]$. For any transfer function matrix G(s), the notation $G^*(s)$ denotes the complex conjugate transpose of $G(-\bar{s})$ with $\bar{s}$ denotes the complex conjugate of s. A power spectrum matrix $\Phi(s)$ can be defined along with two further transfer function matrices W(s) and Z(s) related to $\Phi(s)$ by

$$\Phi(s) = V(s)V^*(s) = W^*(s)W(s) = Z(s) + Z^*(s) \qquad (7.2.2)$$

with W(s) stable and minimum phase and Z(s) the stable part of $\Phi(s)$. Notice that W(s) is unique to within left multiplication by an orthogonal matrix and Z(s) is unique to within addition of a real skew matrix. The functions V(s), W(s) and Z(s) together with a function $F_c(s)$ have the forms

$$\begin{bmatrix} Z(s) & V(s) \\ \\ W(s) & F_c(s) \end{bmatrix} = \begin{bmatrix} D_z & D_v \\ \\ D_w & 0 \end{bmatrix} + \begin{bmatrix} H \\ \\ C \end{bmatrix} (sI-A)^{-1}[G \;\; B] \qquad (7.2.3)$$

where $D_w = D_v^T$, $D_z + D_z^T = D_v D_v^T = D_w^T D_w$, and C and G can be calculated as follows. Define P as the solution of the Lyapunov equation

$$AP + PA^T + BB^T = 0. \qquad (7.2.4)$$

Then

$$G = PH^T + BD_v^T. \qquad (7.2.5)$$

Next, let Q be the stabilizing solution of the Riccati equation

$$QA + A^TQ + (H^T - QG)D_v^{-T}D_v^{-1}(H - G^TQ) = 0 \qquad (7.2.6)$$

and then

$$C = D_v^{-1}(H - G^TQ). \qquad (7.2.7)$$

Notice that as a result,

$$QA + A^TQ + C^TC = 0. \qquad (7.2.8)$$

Definition 7.1: Given the n-th order transfer function matrix V(s), the minimal realizations V(s), W(s) and Z(s) in (7.2.3) are balanced stochastic realizations if

$$P = Q = \Sigma = \mathrm{diag}(\sigma_i) \quad i=1,\ldots,n, \;\; \sigma_i \geq \sigma_{i+1} \qquad (7.2.9)$$

that is, the minimal realization $F_c = C(sI - A)^{-1}B$ in (7.2.3) is internally balanced.

It is of course easy to change the coordinate basis so that balanced stochastic realizations are present. Suppose this is done. Approximation proceeds as follows.

Let Σ be partitioned into two blocks,

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{7.2.10}$$

where $\Sigma_2 = \mathrm{diag}[\sigma_{r+1}, \ldots, \sigma_n]$, and let the other matrices A, B, C, H and G be partitioned correspondingly as

$$\begin{bmatrix} A_r & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \begin{bmatrix} B_r \\ B_2 \end{bmatrix}, [C_r \;\; C_2], \qquad [H_r \;\; H_2], \qquad \begin{bmatrix} G_r \\ G_2 \end{bmatrix} \quad (7.2.11).$$

The BST model-reduction method simply keeps the upper-left corner sub-blocks of all the matrices above to form the reduced transfer functions

$$\begin{bmatrix} Z_r(s) & V_r(s) \\ W_r(s) & F_{cr}(s) \end{bmatrix} = \begin{bmatrix} D_z & D_v \\ D_w & 0 \end{bmatrix} + \begin{bmatrix} H_r \\ C_r \end{bmatrix} (sI - A_r)^{-1}[G_r \;\; B_r] \tag{7.2.12}$$

While $V_r(s)$ in general will not minimize the unweighted version of (7.1.1), it will generally be a good approximant (Harshavardhana *et al.*, 1985, Safonov and Chiang, 1989). Error bound formulas are available in Green (1988a) and Wang and Safonov (1990).

Weighted Balanced Stochastic Truncation

Consider now a stable input-weighting function and associated minimal state-space realization

$$W_I(s) = D_I + C_I(sI - A_I)^{-1}B_I \tag{7.2.13}$$

Let us motivate the approach of the algorithm before presenting the details. The calculations in unweighted BST inter alia compute for V(s) a transfer function $F_c(s)$ which is then reduced by (ordinary additive error) balanced truncation, to yield an approximant $F_{cr}(s)$. Then $V_r(s)$ is found from $F_{cr}(s)$, by reversing the process which yielded $F_c(s)$ from V(s). Evidently, the multiplicative approximation problem is replaced by an additive approximation problem. This is what is done in the weighted case too – and the same weighting is used for the multiplicative problem as for the additive problem.

Proceeding as for the unweighted case, matrices P, G, Q and C are all found. The matrices Q and P are the observability and controllability gramians for the realization {A,B,C} of $F_c(s)$. The basic idea is now to change the controllability gramian of the realization of $F_c(s)$ (equivalently of the realization of V(s)) to reflect the presence of the input weighting $W_I(s)$. The observability gramian Q is unaltered.

The input-weighted transfer function $V(s)W_I(s)$ has a representation with the following state-space matrices

$$\overline{A} = \begin{bmatrix} A & BC_I \\ 0 & A_I \end{bmatrix}, \quad \overline{B} = \begin{bmatrix} BD_I \\ B_I \end{bmatrix}, \quad \overline{C} = [C \;\; DC_I], \quad \overline{D} = DD_I \qquad (7.2.14)$$

and $F_c(s)W_I(s)$ has a realization with the same $\overline{A}$, $\overline{B}$ pair.

Now let

$$\overline{P} = \begin{bmatrix} P & P_{12} \\ P_{12}^T & P_I \end{bmatrix} \qquad (7.2.15)$$

be the solution of the following Lyapunov equation:

$$\overline{P}\,\overline{A}^T + \overline{A}\,\overline{P} + \overline{B}\,\overline{B}^T = 0 \qquad (7.2.16)$$

Assuming there is no pole-zero cancellation in VW_I, $\overline{P}$ is positive definite. So is its principal submatrix P. Now P can be regarded as the frequency weighted controllability gramian for the original transfer function V(s).

Now introduce the coordinate basis change to {A, B, C, H, G} which makes P of (7.2.15) = Q of (7.2.8) = Σ = diag[σ_1, ..., σ_n], $\sigma_i \geq \sigma_{i+1}$, i=1, ... , n-1. This new realization {A, B, C, H, G} is called an input-weighted balanced stochastic realization. (The coordinate basis change is easy to determine, see Green, 1988b).

Partition {A, B, C, H, G} as in (7.2.11) and Σ as in (7.2.10). Now, as previously, the reduction is achieved by eliminating the rows and columns of A, B and H corresponding to Σ_2 in Σ. The reduced order transfer functions of order r are given by (7.2.12), which includes of course $V_r(s)$.

7.3 Example

In this section we present an example to illustrate how a reduced order transfer function obtained via weighted balanced stochastic truncation differs from that obtained in the non-weighted case.

We consider a stable transfer function with two unstable zeros at $z_{o1} = 3.5$ and $z_{o2} = 4$ given by

$$V(s) = (s+1.5)^2(s+2.5)(s-z_{o1})(s-z_{o2})/(s+1)^2(s+2)^2(s+3).$$

In the non-weighted case, the reduced order transfer functions of order 2, 3 and 4 are

$$V_2(s) = (s-z_{o1})(s-z_{o2})/(s^2+3.48s+2.1054)$$

$$V_3(s) = (s+1.4629)(s-z_{o1})(s-z_{o2})/(s^3+4.9618s^2+7.1315s+3.1198)$$

$$V_4(s) \\ = (s+1.1765)(s+3.146)(s-z_{o1})(s-z_{o2})/(s^4+7.8226s^3+20.83s^2+21.855s+7.8966).$$

(It is a standard result that nonminimum phase zeros are preserved, see Green, 1988b). Now we introduce an input-weighting function $W_I(s) = 1/(s+0.1)^2$. Then the reduced order transfer functions are

$$V_{W2}(s) = (s-z_{21})(s-z_{22})/(s^2+3.4168s+2.1323)$$

$$V_{W3}(s) \\ = (s+1.2952)(s-z_{31})(s-z_{32})/(s^3+4.7916s^2+6.5542s+2.7630)$$

$$V_{W4}(s) \\ = (s+1.2294)(s+4.0225)(s-z_{41})(s-z_{42})/(s^4+8.7521s^3+25.324s^2+28.084s+10.55)$$

where

$$z_{21}-z_{o1} = 1.2093\times10^{-2},$$

$$z_{22}-z_{o2} = -1.4849\times10^{-2},$$

$$z_{31}-z_{o1} = 3.7653\times10^{-6},$$

$$z_{32}-z_{o2} = -4.7569\times 10^{-6},$$

$$z_{41}-z_{o1} = 3.4066\times 10^{-8},$$

$$z_{42}-z_{o2} = -4.7531\times 10^{-8}.$$

This fact shows that the number of unstable zeros of the reduced order transfer function is equal to that of the original transfer function, but the locations of unstable zeros are slightly changed.

Now let us define two approximation error functions as

$$e_i(\omega) = |V^{-1}(j\omega)[V(j\omega) - V_i(j\omega)]|, \quad i=2,3,4$$

$$e_{wi}(\omega) = |V^{-1}(j\omega)[V(j\omega) - V_{wi}(j\omega)]|, \quad i=2,3,4.$$

Figs. 7.1, 7.2 and 7.3 show the plots of $\{e_i(\omega),e_{wi}(\omega)\}$ with respect to $\omega\in[10^{-6},10^{2}]$ for i=2,3,4. These figures show that the weighted balanced stochastic truncation method can approximate a given transfer function better in some frequency range than the non-weighted balanced stochastic truncation method, if the weighting function is properly chosen. Of course, at the frequencies where the weighting function has small amplitude, the error is higher, as expected.

7.4 Conclusions

An algorithm for frequency weighted balanced stochastic truncation was suggested and its performance compared with the non-weighted case via an example. The example illustrates the fact that the reduced order transfer functions obtained with the frequency weighted BST method can approximate the original transfer function much better than a reduced order transfer function with non-weighted BST at certain frequencies which are dependent on the weighting. This example also showed that the number of the unstable zeros of the reduced order transfer functions was the same as that of the original transfer function but the zero locations were slightly changed.

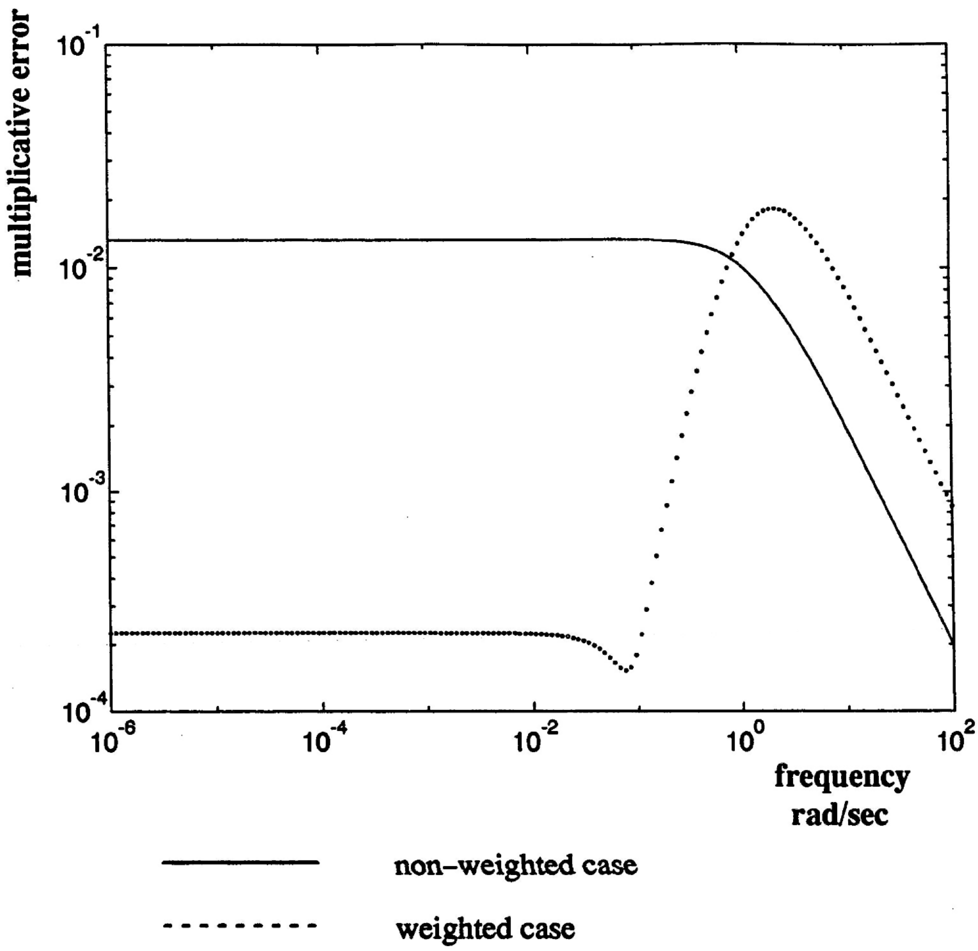


Figure 7.1: Plots of Approximation Errors:
2nd Order Approximations

Issues of interest which remain open are the calculation of error bound formulas and a proof of the preservation of the number of the unstable zeros. One could also examine weighted multiplicative Hankel norm order reduction, generalizing the unweighted ideas of Glover (1986) and Matson *et al.* (1993).

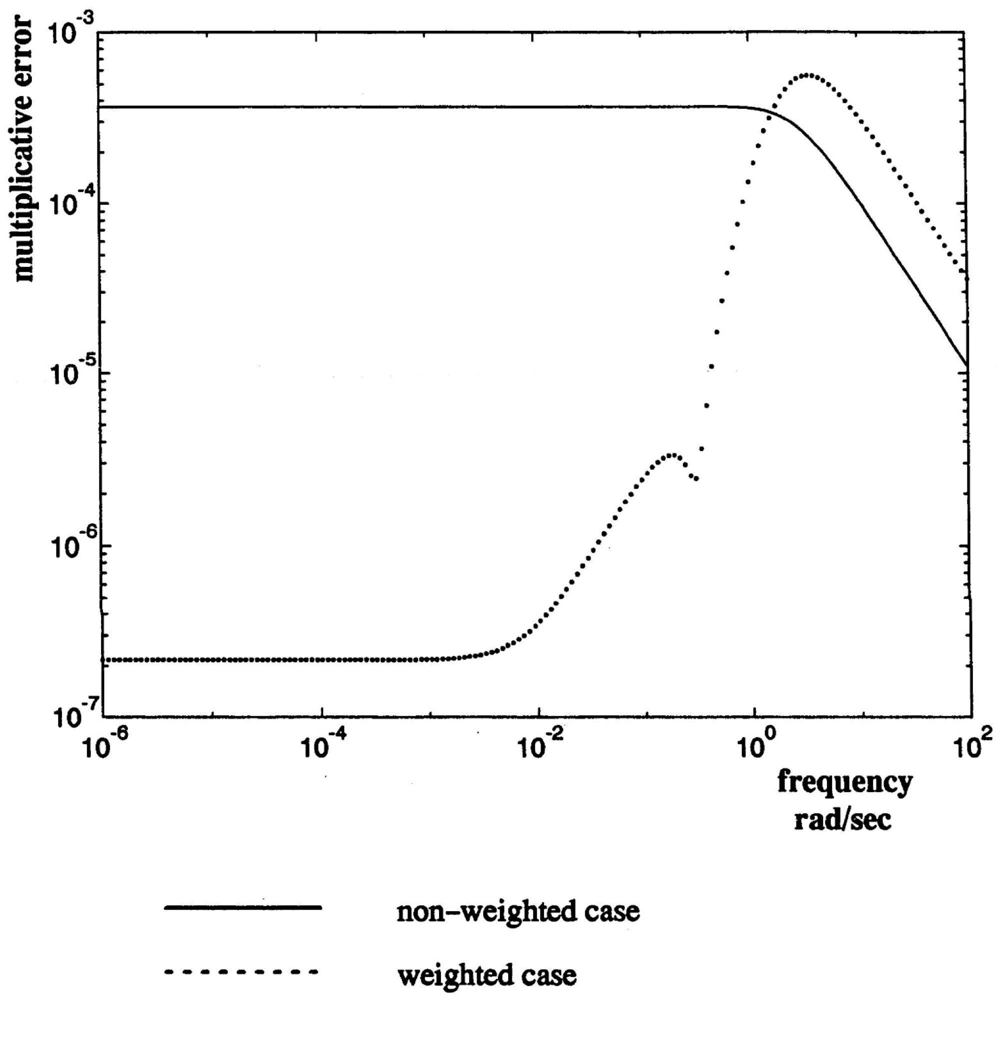


Figure 7.2: Plots of Approximation Errors:
3rd Order Approximations

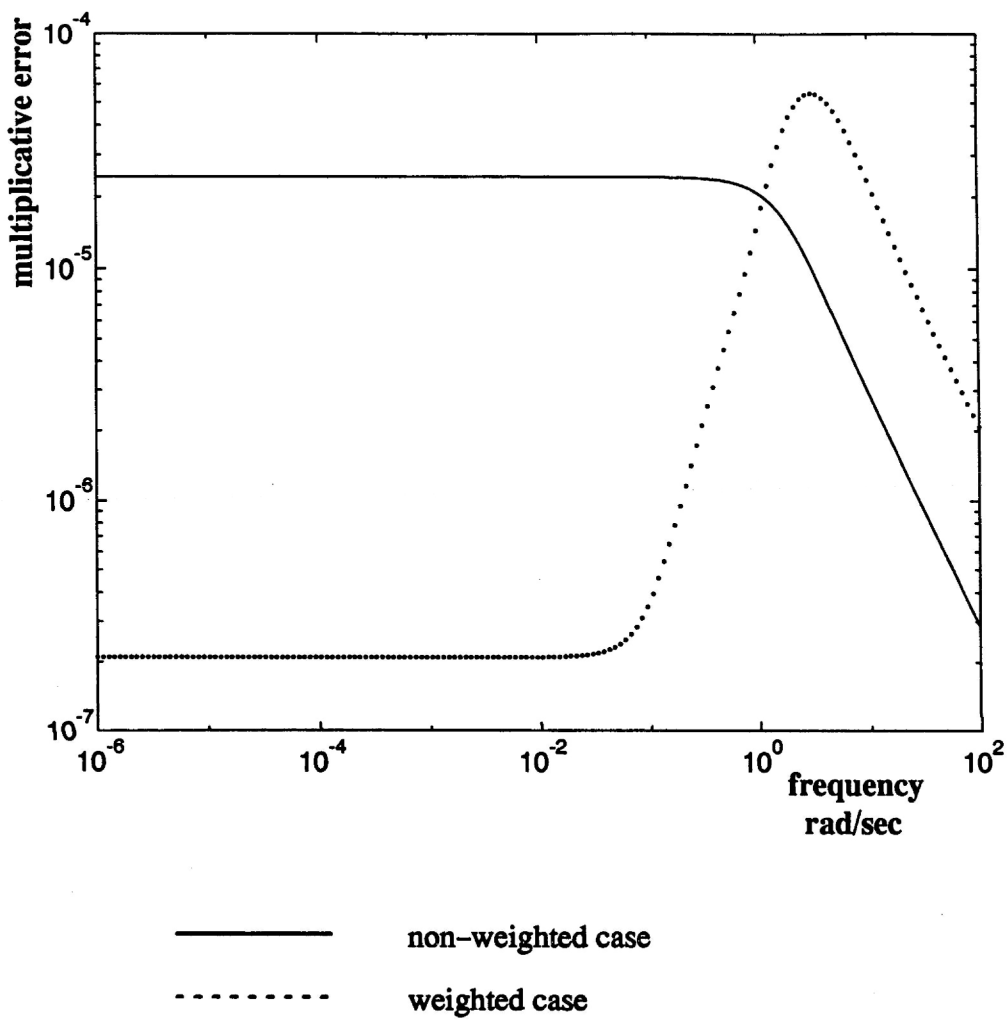


Figure 7.3: Plots of Approximation Errors:
4th Order Approximations

Chapter 8

Conclusions

8.1 Overview of the Thesis

This thesis has been concerned with a number of problems of controller implementation primarily concentrating on sampled–data control systems and on controller order reduction.

We began our investigation with presenting a new approach to establish a frequency–domain paradigm for the sampled–data control systems. The key idea was to associate a sampled–data system with a discrete system which is obtained from the original sampled–data system by very fast sampling at a multiple of the sampling frequency followed by lifting to obtain a single–rate system.

The proposed approach takes inter–sample behaviour of the system into account. Unlike all other approaches based on continuous–time considerations, our approach is essentially discrete–time, but, nevertheless, describes sampled–data systems completely, but with a controllable level of approximation.

We showed that the approach is related to current approaches of Yamamoto and of Araki, Hagiwara and Ito. A number of interesting general theoretical results were established and several typical examples were studied and analyzed.

The problem of an optimal finite word length state-space realization of a digital controller was investigated next. The typical sampled-data closed-loop system was considered consisted of a continuous-time plant, a discrete-time controller, a sampler, a zero-order hold and an antialiasing filter. The definitions of sensitivity "functions" (operators) and L_2 sensitivity measure of a closed-loop system were established. Subsequently, we studied the finite-wordlength-optimal realization minimizing a measure of the sensitivity of the closed-loop operation with respect to controller coefficient errors. The existence and uniqueness of an optimal solution were established. An effective algorithm was proposed to find the optimal sampled-data controller realization minimizing the sensitivity of the closed-loop performance with respect to coefficient errors in the state variable matrices of the controller realization. Studies of two numerical examples confirmed theoretical results.

We continued with investigation of the problem of controller order reduction aimed at preserving the closed-loop performance of a sampled-data system. The fast sampling and lifting procedure allowed capturing of the system's inter-sample behaviour and yielded a time-invariant single-rate system; this then permitted standard order reduction ideas to be applied. Special weighting functions aimed at preserving the closed-loop transfer function were derived and weighted balanced truncation was used to reduce the controller. A practical example confirmed the feasibility of the method.

An error bound for transfer function order reduction was derived next, when frequency weighted balanced truncation was the order reduction method. The bound is valid for both one-sided (input or output) and two-sided weighted balancing approximations with stable weights, which can otherwise be arbitrary. Examples illustrating tightness of the bound were studied.

Next, we showed by examples that a frequency-weighted balanced reduction procedure when applied to a scalar stable transfer function with input and output weights may give an unstable reduced order model or may not give any reduced order model of a particular order. We then proposed a new frequency weighted balanced truncation technique which was guaranteed to yield stable reduced order models even when both input and output weightings were included. The proposed method was essentially a generalisation of Lin and Chiu's (1992) technique and could handle weighting transfer functions which were proper rather than strictly proper. Moreover, a frequency response error bound for the proposed technique was also derived which was applicable to proper (including strictly proper) weighting functions. Comparative analysis of the proposed scheme and previously known techniques was undertaken on several examples.

In the final part of the thesis we generalized the non-weighted balanced stochastic truncation algorithm extending it to allow for frequency weighting. An example illustrated use of the algorithm to secure smaller dB error in selected frequency bands through the introduction of the weighting.

8.2 Further Research

Whilst not an exhaustive list, we propose the following topics for future research that may serve to enhance and extend the research described in this thesis.

- **Appropriate choice of the fast-sampling coefficient N.** It is clear that the larger is N, the better description of the inter-sample behaviour of the sampled-data system we have, with complete description achieved for infinite N. It is also intuitively clear (and results of Shannon, 1949 and Kotelnikov, 1933 support our intuition) that in practice we do not have to take very large N with modest values of N providing usually good approximation and satisfactory system performance. We believe that further research on the simple rules to choose minimal N which guarantees desired performance is required.

- **Roundoff finite wordlength error propagation in sampled-data control systems.** This thesis studies only sensitivity of the sampled-data control system transfer function to the coefficient errors caused by finite wordlength state-space realization. The other source of finite wordlength errors is the roundoff noise gain and it deserves attention to be paid to it.

- **Class of frequency-weighted balanced truncations leading to unstable reduced order controllers.** We showed in the thesis that the double-sided frequency–weighted balanced truncation may yield an unstable reduced order controller which could be very harmful in many applications. A complete description of the class of "unstable" reductions with pre–reduction tests detecting possible stability problems would be very helpful to have and the problem requires, in our opinion, further investigation.

- **Tightening of error bounds for frequency–weighted balanced truncation.** Although we can predict, in the case of the frequency–weighted balanced truncation order reduction method, whether the bound on the error incurred by the order reduction is tight or loose, this knowledge is uncontrollable and does not allow us to improve the bounds. Further research on tightening of the bound and the error bound for the Lin and Chiu's technique generalization may prove to be successful.

- **Extension of the frequency–weighted balanced truncation error bound formula on discrete–time systems.** Such an extension is not automatic, for the correspondingly arising discrete–time Lyapunov equations have more difficult structure, compared to continuous–time equations. Nevertheless, we believe the problem is solvable and is especially important, for it would allow one, by combining the fast–sampling and lifting method with subsequently applied discrete–time case error bound, to handle sampled–data systems providing an error bound for them.

- **Error bound formula for frequency–weighted balanced stochastic truncation.** The advantages of having an *a priori* error bound in the case of frequency–weighted balanced truncation were discussed mostly in Chapter 5 of this thesis. The frequency–weighted balanced stochastic truncation method presented in Chapter 7 lacks an upper error bound and an approach similar to the one used in Chapter 5 may prove to be applicable to this multiplicative order reduction method.

- **Preservation of the number of unstable zeros for frequency–weighted balanced stochastic truncation.** Examples showed that the number of unstable zeros of the reduced order transfer functions was the same as that of the original full order transfer function when the frequency–weighted balanced stochastic truncation scheme was applied. Further investigations are needed in order to prove the fact of number of unstable zeros preservation or, possibly, to construct a counter–example and, further, describe the class of reductions preserving the number of unstable zeros.

- **Weighted multiplicative Hankel norm order reduction.** Glover (1986) and Matson *et al.* (1993) studied unweighted multiplicative order reduction minimizing the corresponding Hankel norm. In Chapter 7 we extended the multiplicative balanced truncation on the frequency–weighted case. Similar ideas may be applied to allow for frequency weighting in multiplicative Hankel norm order reduction.

Bibliography

ACKERMANN, J. (1985). *Sampled-Data Control Systems.* Berlin, Germany: Springer-Verlag.

AL-SAGGAF, U.M. and G.F. FRANKLIN (1988). Model Reduction via Balanced Realization: An Extension and Frequency Weighted Techniques. *IEEE Trans. Automat. Contr.,* AC-33, pp.687-692.

ANDERSON, B.D.O., M.GEVERS and G.LI (1992). Optimal FWL design of a digital controller. *Proc. Eli Jury 70th Birthday Symposium,* Chicago.

ANDERSON, B.D.O. and J.P. KELLER (1994). *Discretization Techniques in Control Systems, Control and Dynamic Systems.* Academic Press, to appear.

ANDERSON, B.D.O. and Y.LIU (1989). Controller Reduction: Concepts and Approaches. *IEEE Trans. Automat. Contr.,* Vol.34, pp.802-812.

ANDERSON, B.D.O. and J.B.MOORE (1971). *Linear Optimal Control.* Englewood Cliffs, New Jersey: Prentice-Hall.

ANDERSON, B.D.O. and J.B.MOORE (1989). *Optimal Control: Linear Quadratic Methods.* Englewood Cliffs, New Jersey: Prentice-Hall.

ARAKI, M., T. HAGIWARA and Y. ITO (1993). Frequency-response of sampled-data systems II: Closed-loop consideration. *12th IFAC World Congress,* Sydney, vol. 7, pp. 293-296.

ARAKI, M., T. HAGIWARA and Y. ITO (1994). Frequency-domain theory of sampled-data systems II: Solution of feedback equations and evaluation of control systems. *IEEE Trans. Automat. Contr.,* to appear.

ARAKI, M. and Y. ITO (1992). On frequency response of sampled-data systems. *Proc. the 21st SICE Symposium on Control Theory,* Kariya, pp. 19–24.

ARAKI, M. and Y. ITO (1993). Frequency-response of sampled-data systems I: Open-loop consideration. *12th IFAC World Congress,* Sydney, vol. 7, pp. 289–292.

ARAKI, M. and Y. ITO (1994). Frequency-domain theory of sampled-data systems I: Definition and properties of FR-operators. *IEEE Trans. Automat. Contr.,* to appear.

ÅSTRÖM, K.J. and B.WITTENMARK (1990). *Computer-Controlled Systems: Theory and Design.* Englewood Cliffs, New Jersey: Prentice Hall.

BAMIEH, B.A. and J.B. PEARSON (1992). A general framework for linear periodic systems with applications to H_∞ sampled-data control. *IEEE Trans. Automat. Contr.,* vol. AC-37, pp. 418–435.

BREWER, J.W. (1978). Kronecker products and matrix calculus in system theory. *IEEE Trans. Circuits Syst.,* Vol. CAS-25, pp. 772–781.

CHEN, T. and B. FRANCIS (1989). Stability of Sampled-Data Systems. *Systems Control Group Report No. 8905,* Dept. of Electrical Engineering, University of Toronto.

CHEN, T. and B. FRANCIS (1990). On the L_2-induced norm of a sampled-data system. *Systems & Control Letters,* vol. 15, pp. 211-219.

CHONGSRID, K. and S.HARA (1994). Stable Controller Reduction for Sampled-Data Systems. *Proc. 33rd IEEE Conf. Decision Contr.,* Orlando, FL, pp. 2708-2713.

DESAI, U.B. and D. PAL (1984). A transformation approach to stochastic model reduction. *IEEE Trans. Automat. Control,* Vol.29, pp. 1097–1100.

DESOER, C.A. and M. VIDYASAGAR (1975). *Feedback systems: input-output properties.* Academic Press, New York.

ENNS, D.F. (1984a). Model Reduction with Balanced Realizations: An Error Bound and a FrequencyWeighted Generalization. *Proc. 23rd IEEE Conf. Decision Contr.*, Las Vegas, NV, pp. 127–132.

ENNS, D.F. (1984b). Model Reduction for Control Systems Design. Ph.D. dissertation, Dept. Aeronautics, Stanford University, Stanford CA.

FRANCIS, B.A. and T.T. GEORGIOU (1988). Stability theory for linear time-invariant plants with periodic digital controllers. *IEEE Trans. Automat. Contr.*, Vol. 33, pp. 820–832.

FRANKLIN, G.F., J.D. POWELL AND M.L. WORKMAN (1990). *Digital Control of Dynamic Systems*. Reading, MA: Addison-Wesley.

FRIEDLAND, B. (1960). Sampled-data control systems containing periodically varying members. *Proc. 1st IFAC World Congress*, Moscow, pp. 361–368.

GEVERS, M. and G.LI (1993). *Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects*. London, UK: Springer Verlag, Communications and Control Engineering Series.

GLOVER, K. (1984). All Optimal Hankel-norm Approximations of Linear Multivariable Systems and their L^∞ -error Bounds. *Int. J. Contr.*, Vol.39, pp.1115–1193.

GLOVER, K. (1986). Multiplicative approximation of linear multivariable systems with L^∞ error bounds, *Proc. Amer. Contr. Conf.*, Seattle, W.A., pp. 1705–1709.

GOODWIN, G.C. and M. SALGADO (1992). Frequency domain sensitivity functions for continuous time systems under sampled data control. Tech. Rep., Dept. Electrical & Computer Engineering, Newcastle Univ.

GREEN, M. (1988a). A relative-error bound for balanced stochastic truncation. *IEEE Trans. Automat. Control*, Vol. 33, pp. 961–965.

GREEN, M. (1988b). Balanced stochastic realization. *Linear Algebra Appl.*, Vol. 98, pp. 211–247.

GREEN, M. and B.D.O. ANDERSON (1990). Generalized balanced stochastic truncation. *Proc. of the 29th IEEE Conf. on Decision Contr.*, Honolulu, pp. 476–481.

HAGIWARA, T., Y. ITO and M. ARAKI (1993). FR-operator and induced norm of sampled-data systems. *Proc. 22nd SICE Symposium on Control Theory,* Kariya, pp. 1-6.

HARA, S., B.D.O. ANDERSON and H. FUJIOKA (1994). Relating H_2 and H_∞ Norm Bounds for Hybrid Systems. *Proc. 33rd IEEE Conf. Decision Contr.*, pp. 724-729.

HARA, S. and P.T. KABAMBA (1990). Worst case analysis and design of sampled-data control systems. *Proc. 29th IEEE Conf. Decision Contr.*, pp. 202-203.

HARSHAVARDHANA, P., E.A. JANCHEERE, and L.M. SILVERMAN (1985). Stochastic balancing and approximation - stability and minimality. *IEEE Trans. Automat. Contr.*, Vol. AC-29, pp. 744-746.

HAYAKAWA, Y., Y. YAMAMOTO and S. HARA (1992). $H\infty$ type problem for sampled-data control systems - A solution via minimum energy characterization. *Proc. 31st IEEE Conf. Decision Contr.*, pp. 463-468.

HELMKE, U. and J.B.MOORE (1995). L^2 sensitivity minimization of linear system representations via gradient flows. *J. Math. Syst., Estimation and Control,* to appear.

HWANG, S.Y. (1977). Minimum Uncorrelated Unit Noise in State-Space Digital Filtering. *IEEE Trans. on Acoust., Speech, and Signal Processing,* Vol. ASSP-25, No. 4, pp. 273-281.

INTEGRATED SYSTEMS INC. (1991). MATRIX_X Model Reduction Module.

ITO, Y., T. HAGIWARA and M. ARAKI (1993). $H\infty$ problem of sampled-data systems viewed from FR-operators. *Proc. 22nd SICE Symposium on Control Theory,* Kariya, pp. 7-12.

JURY, E.I. (1958). *Sampled-Data Control Systems.* New York: John Wiley.

JURY, E.I. (1964). *Theory and Application of the z-Transform Method,* Robert E. Krieger, Huntington, NY.

KABAMBA, P. and S. HARA (1990). On computing the induced norm of sampled-data systems. *Proc. of American Control Conf.*, pp. 319-320.

KELLER, J.P. and B.D.O.ANDERSON (1991). H∞-Optimal Controller Discretization. *Int. J. of Robust and Nonlinear Control,* Vol.1, p.125–137.

KELLER, J.P. and B.D.O.ANDERSON (1992). A new approach to the discretization of continuous-time controllers. *IEEE Trans. Automat. Contr.,* Vol. 37, pp. 214–223.

KHARGONEKAR, P.P., K.POOLLA and A.TANNENBAUM (1985). Robust Control of Linear Time-Invariant Plants Using Periodic Compensation. *IEEE Trans. Automat. Contr.,* Vol. AC-30, p.1088–1096.

KIM, S.W., B.D.O.ANDERSON and A.G.MADIEVSKI (1995). Error Bound for Transfer Function Order Reduction Using Frequency Weighted Balanced Truncation. *Syst. Contr. Lett.,* Vol.24, issue 3, to appear.

KOTELNIKOV, V.A. (1933). On the Transmission Capacity of 'Ether' and Wire in Electrocommunications. *Proc. First All-union Conference on Questions of Communication.,* Moscow, USSR.

LATHAM, G.A., and B.D.O.ANDERSON (1985). Frequency-Weighted Optimal Hankel-norm Approximation of Stable Transfer Functions. *Syst. Contr. Lett.,* Vol.5, p.229–236.

LI, G. (1990). Finite Precision aspects in the parametrization of control, estimation, and filtering problems. Ph.D. thesis, Laboratoire d'automatique, de dynamique et d'analyse des systemes, Universite Catholique De Louvain, Louvain-La-Neuve, Belgium.

LI, G., B.D.O.ANDERSON, M.GEVERS and J.E.PERKINS (1992). Optimal FWL design of state-space digital systems with weighted sensitivity minimization and sparseness consideration. *IEEE Trans. Circuits Syst.,* Vol. 39, pp. 365-377.

LI, G. and M.GEVERS (1990a). Sensitivity and Roundoff Noise Optimization of a State-Estimate Feedback Controller. *Proc. 11th IFAC World Congress,* Tallinn, USSR, Vol. 4, pp. 303–310.

LI, G. and M.GEVERS (1990b). Optimal Finite Precision Implementation of a State-Estimate Feedback Controller. *IEEE Trans. on Circuits and Systems,* Vol. CAS-38, No. 12, pp. 1487–1498.

LI, G. and M.GEVERS (1991). Minimization of Finite Wordlength Effects in Compensator Design. *Proc. of the 1st European Control Conference,* Grenoble, France, pp. 544–549.

LI, G. and M.GEVERS (1993). Compensator Realizations that Minimize the Closed Loop Pole Sensitivity. *Proc. 12th IFAC World Congress*, Sydney, Australia, Vol. 5, pp. 149–152.

LIN, C.-A. and T.-Y. CHIU (1992). Model Reduction via Frequency Weighted Balanced Realization. *Control Theory and Advanced Technology*, Vol. 8, pp. 341–451.

LIU, K. and R.SKELTON (1990). Optimal Controllers for Finite Wordlength Implementation. *Proc. of the American Control Conference*, San Diego, pp.1935–1940.

LIU, K., R.SKELTON and K.GRIGORIADIS (1992). Optimal Controllers for Finite Wordlength Implementation. *IEEE Trans. Automat. Contr.*, Vol. 37, No. 9, pp. 1294–1304.

MATSON, J.B., L. LAM, B.D.O. ANDERSON and B. JAMES (1993). Multiplicative Hankel norm approximation of multivariable systems. *Int. J. Contr.*, Vol. 58, pp. 129–167

MCFARLANE, D.C. and K.GLOVER (1990). *Robust Controller Design Using Normalized Coprime Factor Plant Description*. Berlin: Springer–Verlag.

MILNOR, J. (1963). *Morse Theory*. Princeton, New Jersey: Princeton University Press.

MOORE, B.C. (1981). Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction. *IEEE Trans. Automat. Contr.*, Vol. AC–26, pp.17–32.

MORI, T., N.FUKUMA and M.KUWAHARA (1986). Explicit Solution and Eigenvalue Bounds in the Lyapunov Matrix Equation. *IEEE Trans. Automat. Contr.*, Vol. 31, pp. 656–658.

MULLIS, C.T. and R.A.ROBERTS (1976). Synthesis of Minimum Roundoff Noise Fixed–Point Digital Filters. *IEEE Trans. on Circuits and Systems*, Vol. CAS–23, pp. 551–562.

NEUDECKER, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Stat. Assoc.*, Vol. 64, pp.

953–963.

PEARSON, J.B., B.A. FRANCIS and A. TANNENBAUM (1991). A lifting technique for linear periodic systems with applications to sampled-data control. *Systems & Control Letters*, vol. 17, pp. 79–88.

PERKINS, J.E., U.HELMKE and J.B.MOORE (1990). Balanced realizations via gradient flow techniques. *Syst. Contr. Lett.*, Vol. 14, pp. 369–380.

PERNEBO, L. and L.M.SILVERMAN (1982). Model Reduction via Balanced State Space Representations. *IEEE Trans. Automat. Contr.*, Vol. 27, pp. 382–387.

PERRY, T. (1989). M.A.Sc. thesis, Dept. of Elec. Eng., Univ. of Toronto, Toronto, Canada.

SAFONOV, M.G., R.Y. CHIANG and H. FLASHNER (1988). H_∞ robust control synthesis for a large space structure. *Proc. of American Control Conf.*, Atlanta, GA, pp. 2038–2045.

SAFONOV, M.G. and R.Y. CHIANG (1989). Model reduction for robust control: A Schur relative-error method. *Int. J. Adaptive Control Signal Process*, Vol.2, pp. 252–272.

SALEHI, S. (1985). Application of Adaptive Observers to the Control of Flexible Spacecraft. *10^{th} IFAC Symposium, "Automatic Control in Space"*, Toulouse.

SHANNON, C.E. (1949). Communication in Presence of Noise. *Proc. IRE.*, Vol. 37, pp. 10–21.

SIVASHANKAR, N. and P.P. KHARGONEKAR (1992). Worst case performance analysis of linear systems with jumps with application to sampled-data systems. *Proc. of American Control Conf.*, pp. 692–696.

SREERAM, V., B.D.O. ANDERSON and A.G. MADIEVSKI (1994). New Results on Frequency Weighted Balanced Reduction Technique. *IEEE Trans. Automat. Contr.*, submitted.

SUN, W., K.M. NAGPAL and P.P. KHARGONEKAR (1991). H∞ control and filtering with sampled measurements. *Proc. of American Control Conf.*, pp. 1652–1657.

TADMOR, G. (1992). H∞ optimal sampled-data control in continuous time systems. *Int. J. Control*, vol. 56, pp. 99–141.

THIELE, L. (1986). On the sensitivity of linear state-space systems. *IEEE Trans. Circuits Syst.*, Vol. CAS-33, pp. 502-510.

TOIVONEN, H.T. (1992). Sampled-data control of continuous-time systems with an H∞ optimality criterion. *Automatica*, vol. 28, pp. 45-54.

TROCH, I. (1987). Improved Bounds for the Eigenvalues of Solutions of Lyapunov Equations. *IEEE Trans. Automat. Contr.*, Vol. 32, pp. 744-747.

WAGIE, D.A. and R.E. SKELTON (1986). A Projection Approach to Covariance Equivalent Realization of Discrete Systems. *IEEE Trans. on Automat. Contr.*, AC-31, pp. 1114-1120.

WANG, W. and M.G. SAFONOV (1990). A tighter relative-error bound for balanced stochastic truncation. *Systems and Control Letters*, Vol. 14, pp. 307-317.

WILLIAMSON, D. (1991). *Digital Control and Implementation: Finite Wordlength Considerations*. New York, Sydney: Prentice Hall.

WILLIAMSON, D. and K.KADIMAN (1989). Optimal Finite Wordlength Linear Quadratic Regulation. *IEEE Trans. on Automat. Contr.*, Vol. AC-34, No. 12, pp. 1218-1228.

YAMAMOTO, Y. (1993). On the state space and frequency domain characterization of H_∞-norm of sampled-data systems. *Systems & Control Letters*, vol. 21, pp. 163-172.

YAMAMOTO, Y. (1994). A function space approach to sampled-data control systems and tracking problems. *IEEE Trans. Automat. Contr.*, vol. AC-39, pp. 703-712.

YAMAMOTO, Y. and M. ARAKI (1994). Frequency responses for sampled-data systems - Their equivalence and relationships. *Linear Algebra and Its Applications*, vols. 205-206, pp. 1319-1339.

YAMAMOTO, Y. and P.P. KHARGONEKAR (1993). On the frequency response of sampled-data systems. *Proc. 32nd IEEE Conf. Decision Contr.*, San Antonio, pp. 799-804.

ZHANG, C. and J. ZHANG (1994). Orthonormal Lifting for Digital Control Systems. *Proc. ASCC*, vol. 3, pp. 73-76.

Someone told me that each equation
I included in the book would halve
the sales.

Stephen Hawking *A Brief History of Time*

Appendix A

Proof of Lemma 2.1

(i) Follows by Desoer and Vidyasagar (1975), p.113.

(ii) Let $u(\cdot)\in L_p(-\infty,\infty)$. Consider

$$y(t+\delta)-y(t) = \int_{-\infty}^{t+\delta} T(t+\delta,s)\, u(s)\, ds - \int_{-\infty}^{t} T(t,s)\, u(s)\, ds$$

$$= \int_{-\infty}^{t_0} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds + \int_{t_0}^{t} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds + \int_{t}^{t+\delta} T(t+\delta,s)\, u(s)\, ds$$

for any $t_0<t$.

Now

$$\left\| \int_{-\infty}^{t_0} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds \right\|$$

$$\leq \alpha \int_{-\infty}^{t_0} \{\exp[-\beta(t+\delta-t_0) + \exp[-\beta(t-t_0)]\}\exp[-\beta(t_0-s)]\, \|u(s)\|\, ds$$

$$\leq 2\alpha \exp[-\beta(t-t_0)] \int_{-\infty}^{t_0} \exp[-\beta(t_0-s)]\, \|u(s)\|\, ds$$

$$\leq 2\alpha \exp[-\beta(t-t_0)]\, \|u\|_p / \beta$$

(The last inequality follows by Desoer and Vidyasagar (1975), p.113)

Also since $T(\cdot,\cdot)$ is continuous on $t_0 \leq s \leq t$; hence, given arbitrary $\bar{\varepsilon} > 0$, there exists $\bar{\delta}(\bar{\varepsilon},t_0,t)$ such that

$$\|T(t+\delta,s)-T(t,s)\| \leq \bar{\varepsilon} \qquad \forall \quad \delta \leq \bar{\delta},\ t_0 \leq s \leq t$$

Then

$$\| \int_{t_0}^{t} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds \| \leq \bar{\varepsilon}\, \|u\|_p\, (t-t_0)$$

Finally,

$$\| \int_{t}^{t+\delta} T(t+\delta,s)\, u(s)\, ds \| \leq \alpha\, \|u\|_p\, \delta$$

Consequently,

$$\|y(t+\delta)-y(t)\| \leq \{2\alpha \exp[-\beta(t-t_0)] / \beta + \bar{\varepsilon}\, (t-t_0) + \alpha\, \delta\}\, \|u\|_p.$$

Let $\varepsilon>0$. Choose $t_0<t$ so that

$$2\alpha \exp[-\beta(t-t_0)] / \beta < \varepsilon/3.$$

Choose

$$\bar{\varepsilon} = \varepsilon/3(t-t_0).$$

Choose

$$\delta < \min\{\varepsilon/3\alpha,\ \bar{\delta}(\bar{\varepsilon},t_0,t)\}.$$

Then

$$\|y(t+\delta) - y(t)\| < \varepsilon.$$

This establishes (ii).

(iii) Observe that

$$y(t) = \int_{t-\tau}^{t} T(t,s)\, u(s)\, ds + \int_{t-2\tau}^{t-\tau} T(t,s)\, u(s)\, ds + \ldots$$

$$\|y(t)\| \leq \alpha \int_{t-\tau}^{t} \|u(s)\|\, ds + \alpha\, e^{-\beta\tau} \int_{t-2\tau}^{t-\tau} \|u(s)\|\, ds + \ldots$$

$$\leq \alpha/(1-e^{-\beta\tau})\ \sup_k \int_{t-(k+1)\tau}^{t-k\tau} \|u(s)\|\, ds$$

$$\leq \alpha/(1-e^{-\beta\tau})\ \sup_k\ \tau^{1/q} \left[\int_{t-(k+1)\tau}^{t-k\tau} \|u(s)\|^p\, ds\right]^{1/p} \quad (\text{with } 1/p + 1/q$$

$=1)$

$$\leq \alpha\, \tau^{1/q} /(1-e^{-\beta\tau})\ \sup_T \left[\int_{T}^{T+\tau} \|u(s)\|^p\, ds\right]^{1/p}$$

It is then trivial that $y(\cdot) \in L_{pav}$.

To prove the continuity result, suppose $\varepsilon>0$ is given. Write

$$y(t) = \int_{t-K\tau}^{t} T(t,s)\, u(s)\, ds + \int_{-\infty}^{t-K\tau} T(t,s)\, u(s)\, ds = y_1(t) + y_2(t).$$

Choose K sufficiently large that

$$\alpha\, \tau^{1/q}\, e^{-\beta K\tau} /(1-e^{-\beta\tau}) < \varepsilon/4$$

Then $y_1(t)$ is continuous at t since the restriction of $u(\cdot)$ to $[t-K\tau, t+\delta]$ is in $L_p(-\infty,\infty)$. Also

$\|y_2(t+\delta) - y_2(t)\| < \varepsilon/2\ \sup_T \left[\int_{T}^{T+\tau} \|u(s)\|^p\, ds\right]^{1/p}$ because of the way K has been chosen [use a minor variant of the argument above establishing the bound on $\|y(t)\|$] These facts allow us to conclude the existence of $\bar{\delta}$ such that for $0 \leq \delta \leq \bar{\delta}$, $\|y(t+\delta) - y(t)\| < \varepsilon$. Hence $y(\cdot)$ is continuous.

Appendix B

Proof of Corollary 2.1

Let $\varepsilon>0$ be specified. Choose $\bar{\sigma}$ such that

$$2\,\alpha\,\beta^{-1}\exp[-\beta\,\bar{\sigma}\,] < \varepsilon/3$$

and let M be a positive integer such that

$$(M-1)\,\tau \leq \bar{\sigma} < M\,\tau$$

Choose $\hat{\delta}(\varepsilon)$ with $\hat{\delta} < \tau$ so that

$$\|T(\sigma+\delta,\varrho) - T(\sigma,\varrho)\| < \varepsilon/(3\,M\,\tau)$$

$$\forall\,\delta \leq \hat{\delta},\quad 0 \leq \varrho \leq \sigma \leq \sigma + \delta \leq (M+2)\,\tau$$

This can be achieved because of the continuity of $T(t,s)$ in $0 \leq s \leq t \leq (M+2)\,\tau$.

Now let t be arbitrary, and choose t_0 such that $t - t_0 = \bar{\sigma}$.

Choose an integer P such that

$$P\tau \leq t_0 \leq (P+1)\,\tau$$

Then, using the periodicity property, we have

$$T(t+\delta,s) - T(t,s) = T(t+\delta-P\tau,\ s-P\tau) - T(t-P\tau,\ s-P\tau) = T(\sigma+\delta,\ \varrho) - T(\sigma,\ \varrho)$$

where $\varrho = s - P\tau$, $\sigma = t - P\tau$. Observe further that

$$\varrho = s - P\tau \geq s - t_0$$

and

$$\sigma = t - P\tau = (t - t_0) + (t_0 - P\tau) \leq \bar{\sigma} + \tau < (M+1)\,\tau$$

Hence

$$\sigma + \delta < (M+2)\,\tau.$$

If we now make the restriction $t_0 \leq s \leq t \leq t + \delta$, we have $0 \leq \varrho \leq \sigma \leq \sigma + \delta \leq (M + 2)\,\tau$, so that

$$\|T(t+\delta,s) - T(t,s)\| = \|T(\sigma+\delta,\varrho) - T(\sigma,\varrho)\| < \varepsilon/(3M\tau)$$

Then

$$\left\| \int_{t_0}^{t} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds \right\| \leq \varepsilon/(3M\tau)\, \|u\|_p\, (t-t_0) < \varepsilon/3\, \|u\|_p.$$

Further, as shown in the course of proving Lemma 2.1, part (ii),

$$\left\| \int_{-\infty}^{t_0} [T(t+\delta,s)-T(t,s)]\, u(s)\, ds \right\| \leq 2\alpha \exp[-\beta(t-t_0)]\, \|u\|_p\, /\beta.$$

The choice of $\bar{\sigma}$ and the definition of t_0 by $(t-t_0)=\bar{\sigma}$ ensures that the right side is bounded by $\varepsilon/3$. Also, as in proving Lemma 2.1, part (ii),

$$\left\| \int_{t}^{t+\delta} T(t+\delta,s)\, u(s)\, ds \right\| \leq \alpha\, \|u\|_p\, \delta.$$

The result for $L_p(-\infty,\infty)$ then follows by choosing $\bar{\delta} = \min\,[\varepsilon / 3\alpha,\ \hat{\delta}(\varepsilon)]$. The latter part in the proof of Lemma 2.1 part (iii) establishes the result for L_{pav}.

Appendix C

Proof of Lemma 2.2

$H_h\ S_h$ is bounded on $L_\infty \cap C$, see Chen and Francis (1989). This is also easily checked. Now assume $u(\cdot) \in L_1$ with

$$y(t) = \int_{-\infty}^{t} T(t,s)\ u(s)\ ds.$$

Then

$$\sum_{k=-\infty}^{\infty} \|y(t+kh)\| = \Sigma_k \left\| \int_{-\infty}^{\infty} T(t+kh,s)\ u(s)\ ds \right\|$$

$$\leq \Sigma_k \int_{-\infty}^{\infty} \|T(t+kh,s)\|\ \|u(s)\|\ ds$$

$$\leq \int_{-\infty}^{\infty} \left\{ \sum_{k=-\infty}^{\infty} \|T(t+kh,s)\| \right\} \|u(s)\|\ ds.$$

Now

$$\sum_{k=-\infty}^{\infty} \|T(t+kh,s)\| \leq \sum_{k=0}^{\infty} \alpha \exp[-\beta kh] = \alpha/(1-e^{-\beta h}).$$

Hence

$$\sum_{k=-\infty}^{\infty} \|y(t+kh)\| \leq \alpha/(1-e^{-\beta h}) \|u\|_1.$$

Consequently

$$\|H_h S_h T u\| \leq \alpha h/(1-e^{-\beta h}) \|u\|_1.$$

Hence $H_h S_h T$ is a bounded operator on $L_1(-\infty,\infty)$. Being a bounded operator on $L_\infty(-\infty,\infty)$, this ensures it is a bounded operator on $L_p(-\infty,\infty)$.

Note that as $h \to 0$ $\alpha h/(1-e^{-\beta h}) \to \alpha \beta^{-1}$, so no problems arise with vanishingly small h.

(ii) Next, suppose that $u \in L_{pav}(-\infty,\infty)$. We have already seen that

$$\|y(t)\| \leq \alpha h^{1/q} /(1-e^{-\beta h}) \sup_T [\int_T^{T+h} \|u(s)\|^p ds]^{1/p}.$$

It follows that $H_h S_h y(\cdot)$ satisfies the same bound, and so

$$\sup_T [\int_T^{T+h} \|H_h S_h T u\|^p dt]^{1/p}$$

$$\leq \alpha h^{1/q} /(1-e^{-\beta h}) \sup_T [\int_T^{T+h} \|u(s)\|^p ds]^{1/p} [\int_T^{T+h} dt]^{1/p}$$

$$= \alpha h /(1-e^{-\beta h}) \sup_T [\int_T^{T+h} \|u(s)\|^p ds]^{1/p} .$$

As before, no problems arise when $h \to 0$.

Appendix D

Proof of Lemma 2.3

First suppose $u \in L_\infty(-\infty,\infty)$. Let $\varepsilon>0$. By the Corollary,

$$\|[y-H_hS_hy](t)\|_\infty = \|y(t) - y(kh)\|_\infty \qquad kh \le t<(k+1)h$$

$$< \varepsilon \, \|u\|_\infty$$

whenever $h < \bar{\delta}(\varepsilon)$. So the result is immediate.

Next, let $u \in L_1(-\infty,\infty)$, $\varepsilon>0$. Then with $kh \le t < (k+1)h$,

$$\|y - H_hS_hy\|_1 = \sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} \| \int_{-\infty}^{t} T(t,s)\, u(s)\, ds - \int_{-\infty}^{kh} T(kh,s)\, u(s)\, ds \| \, dt.$$

Noting that $T(kh,s)=0$ if $s>kh$, we can rewrite this as

$$\|y - H_hS_hy\|_1 = \sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} dt \, \| \int_{-\infty}^{t} [T(t,s) - T(kh,s)]\, u(s)\, ds \|$$

$$\leq \int_{-\infty}^{t} ds \; \{ \sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} \|T(t,s) - T(kh,s)\| \, dt \} \; \|u(s)\|.$$

We will now show that

$$\sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} \|T(t,s) - T(kh,s)\| \, dt$$

goes to zero as $h \rightarrow 0$. Without loss of generality, because of the periodicity we may suppose that $0 \leq s \leq h$. Choose $\varepsilon > 0$ and t_1 such that

$$\int_{t_1}^{\infty} \|T(t,s)\| dt < \varepsilon/4$$

$$\sum_{kh \geq t_1} h \, \|T(kh,s)\| < \varepsilon/4.$$

Choose h sufficiently small so that

$$\|T(t,s) - T(kh,s)\| < \varepsilon \, / \, 2t_1$$

in the finite region defined by $0 \leq s \leq h$, $0 \leq t \leq t_1+h$, and $kh \leq t < (k+1)h$.

Then with $K = \lceil t_1/h \rceil$,

$$\sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} \|T(t,s) - T(kh,s)\| \, dt$$

$$= \sum_{k \geq K}^{\infty} \int_{kh}^{(k+1)h} \|T(t,s)\| \, dt + \sum_{k \geq K}^{\infty} h \, \|T(kh,s)\| \, dt$$

$$+ \sum_{k=-\infty}^{K-1} \int_{kh}^{(k+1)h} \|T(t,s) - T(kh,s)\| \, dt$$

$$< \varepsilon/4 + \varepsilon/4 + \varepsilon/2 = \varepsilon.$$

It follows that

$$\|y - H_h \, S_h \, y\|_1 \leq \int_{-\infty}^{t} ds \;\; \varepsilon \, \|u(s)\| = \varepsilon \, \|u\|_1.$$

Since the result holds for L_1 and L_∞, it holds for L_p, $1 \leq p \leq \infty$.

The result for L_{pav} is a very simple consequence of Corollary 2.1.

Appendix E

Proof of Theorem 2.1

Let us consider a sampled–data control system shown in Fig. 2.1, where G(s) is a continuous–time plant and K(z) is a discrete–time controller.

Let us also assume that the plant is given by its state–space matrices (A,B,C,D) as in (2.2.1) and the matrix A satisfies the following condition:

Condition E. If μ is an eigenvalue of A with positive real part, none of the points $\mu+2\pi k/\tau$ is an eigenvalue of A.

Obviously, the condition is a mild one for it is satisfied if τ is small enough.

Outline of the proof of Theorem 2.1.
In their 1994 work Anderson and Keller adjusted the definition of norms of discrete–time sequences. Namely, for a sequence a_i in Z_+ obtained with sampling interval h, they defined the norm of the sequence as

$$||a||_p = h^{1/p} [\Sigma_i |a_i|^p]^{1/p}. \tag{E1}$$

This slight change in the traditional definition was motivated by the advantage of having equal p–norms for a discrete–time sequence as input to a hold element and for the hold's continuous–time output.

Working in the space with this adjusted norm definition, Anderson and Keller obtained results analogous to Lemmata 2.1, 2.2 and 2.3. These results were essential in establishing the following lemma:

Lemma E1. (Anderson and Keller, 1994, see Theorem B.1) Let T be the operator defining the sampled–data system of Fig.2.1, for which Condition E holds and the compensator is stabilizing. Then for all $1 \leq p \leq \infty$ there holds

$$\|T\|_p = \lim_{N \to \infty} \|S_{\tau/N}\, T\, H_{\tau/N}\|_p$$

(The norm is meant in the sense of definition E1)

The lifting preserves induced operator norms, for it is just a regrouping of input and output values not affecting the sum of the elements, while the input and output scaling factors do change along with the sampling interval, but cancel out.

Now lifting of the fast–sampled system $S_{\tau/N}\, T\, H_{\tau/N}$ results in the statement of the theorem:

$$\|T\|_p = \lim_{N \to \infty} \|\mathcal{T}_{[N]}\|_p \; .$$

Appendix F

Proof of Theorem 3.2

The proof of the Theorem can be obtained analogously to a proof of a similar result in Gevers and Li (1993) and Helmke and Moore (1991), but requires certain preliminaries. The first lemma is a variation on a standard result for systems with continuous-time inputs and outputs.

Lemma F.1. Consider a periodically time-varying causal linear system S_1 with continuous-time input (of arbitrary dimension) and scalar discrete-time output with sampling interval T, the underlying period of S_1, and a periodically time-varying causal linear system S_2 with scalar discrete-time input with the same sampling interval and period, and with continuous-time output (of arbitrary dimension). Denote the impulse response values of S_1 and S_2 by the row vector $\alpha_1(kT,s)$ and column vector $\alpha_2(t,kT)$, with $k \in Z$, and $s,t \in \mathbb{R}$. Let S_3 denote S_1 followed by S_2, and have impulse response

$$\alpha_3(t,s)=\sum_{t\geq kT\geq s} \alpha_2(t,kT)\ \alpha_1(kT,s) \tag{F1}$$

Then α_3 is the zero impulse response if and only if one at least of α_1, α_2 has this property.

Proof. Suppose neither of α_1, α_2 is the zero impulse response. Choose s so that for some k, $\alpha_1(kT,s)\neq 0$. Let k_1 be the least such k. Because α_2 is not a zero impulse response, and is periodically time-varying, $\alpha_2(t,k_1T)$ is not identically zero as a function of t. Choose $t\in[(m-1)T,mT]$ for which $\alpha_2(t,k_1T)$ is nonzero and m is minimal. Then

$$\alpha_3(t,s)=\alpha_2(t,k_1T)\ \alpha_1(k_1T,s)\neq 0 \tag{F2}$$

This lemma is solely used to establish the next lemma; it will assure us that there is no infimum for $M_2(P)$ involving a singular P, or singular P^{-1}.

Lemma F.2. Under the hypotheses of Theorem 3.2, J_B and J_C are positive definite matrices.

Proof. We shall focus on J_C only, the proof for J_B being comparable. By (3.3.2b), (3.3.4b), J_C is nonnegative definite and is singular if and only if for some nonzero control vector $\alpha=[\alpha_1,\alpha_2,...,\alpha_R]^T$, there holds

$$\partial\mathcal{G}_{k,l}(t,s)/\partial C\ \alpha = 0 \tag{F3}$$

for all t, s, k and l. By (3.2.10), this means that

$$\mathcal{V}^T\ e_k\ e_l^T\ \mathcal{W}_A^T\ \alpha = 0 \tag{F4}$$

for all t, s, k, and l. Now the operator $\mathcal{V}$ is depicted in Fig. 3.2. It is evident that $\mathcal{V}$ is not identically zero, so that for some choice of j and k, $\mathcal{V}_{k,j} = e_k^T\ \mathcal{V}\ e_j = e_j^T\ \mathcal{V}^T\ e_k$ is not a zero impulse response. If (F4) holds, left multiplication by e_j^T yields

$$\mathcal{V}_{k,j}\ e_l^T\ \mathcal{W}_A^T\ \alpha = 0 \tag{F5}$$

and by Lemma F.1, we must have

$$e_l^T\ \mathcal{W}_A^T\ \alpha = 0 \qquad \text{(for all l)} \tag{F6}$$

or

$$\alpha^T\ \mathcal{W}_A\ e_l = 0 \tag{F7}$$

for all l. This means that the impulse response of the system $\alpha^T\ \mathcal{W}_A$ is zero, see Fig. 3.3, where α is a nonzero constant R-vector. The set up is redrawn on Fig. F.1.

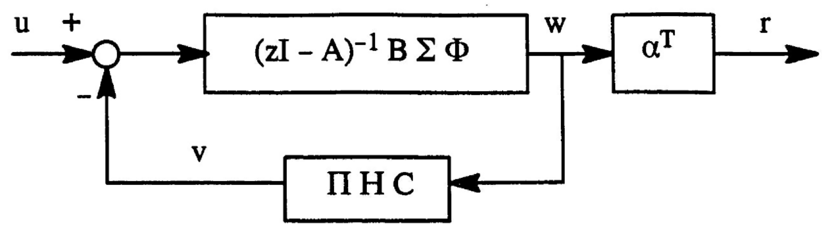


Figure F.1

For all u(·) in Fig. F.1, $r \equiv 0$ results. Consider u(·) in Fig. F.1 generated by the arrangement in Fig. F.2. It follows that for all $\bar{u}$ in Fig. F.2, $r \equiv 0$. This is equivalent to having all $\bar{u}$ in Fig. F.3 produce $r \equiv 0$. This can only happen if $\alpha^T(zI-A)^{-1}B=0$. But since (A,B) is controllable (by the minimality assumption, which is included in the hypothesis of Theorem 3.2), and $\alpha \neq 0$, by assumption above, a contradiction results.

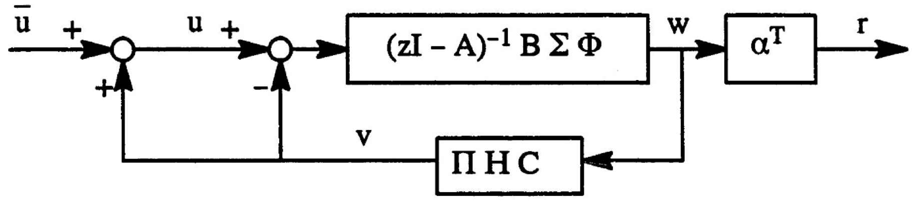


Figure F.2

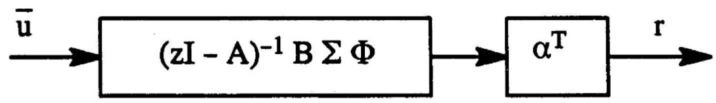


Figure F.3

The proof then follows by an application of Theorem 5.1 of Gevers and Li (1993). Alternatively, following ideas of Helmke and Moore (1991), one can obtain a proof as follows. We use the following idea:

Definition F.1. Let J(P) be an $n \times n$ matrix J. Then the linearization of J(·) at a value of the argument P is that matrix L for which, as $||\Delta P|| \to 0$,

$$\operatorname{vec}\,[J(P+\Delta P)-J(P)]=L\,\operatorname{vec}\,\Delta P + o||\Delta P||. \tag{F8}$$

We shall identify J(P) with $\partial M_2(P)/\partial P$ in (3.3.11). It is then possible to show that for every positive definite P_∞ for which $J(P_\infty) = 0$, the linearization L is positive definite, i.e. every extreme point of $M_2(P)$ is a minimum. The properties of J_B and J_C assure that $M_2(P)$ assumes its global minimum. Morse theory, see e.g. Milnor (1963) (and as explained in Helmke and Moore (1991)) ensures that there exists a single local minimum, viz. the global minimum, of $M_2(P)$. (Let us consider a smooth real function m of a real argument p, such that $\lim_{p \to \infty} m(p) = +\infty$ and $\lim_{p \to -\infty} m(p) = +\infty$. If all the extremum points of m(p) are minima, then m(p) has a unique minimum. Morse theory is a natural extension of this idea to the multidimensional case).

Appendix G

Proof of Lemma 3.3

By definition

$$J_B = \int_0^{\tau} dt \int_{-\infty}^{t} (\partial \mathcal{H}_{k,l}(t,s)/\partial B)\ (\partial \mathcal{H}_{k,l}(t,s)/\partial B)^T ds$$

Hence

$$= \int_0^{\tau} dt \int_{-\infty}^{t} \begin{pmatrix} h_{B\,11}(t,s) \dots h_{B\,1L}(t,s) \\ \vdots \qquad\qquad \vdots \\ h_{B\,R1}(t,s) \dots h_{B\,RL}(t,s) \end{pmatrix} \begin{pmatrix} h_{B\,11}(t,s) \dots h_{B\,R1}(t,s) \\ \vdots \qquad\qquad \vdots \\ h_{B\,1L}(t,s) \dots h_{B\,RL}(t,s) \end{pmatrix} ds$$

and

$$(J_B)_{i,j} = \sum_{m=1}^{L} \int_0^{\tau} dt \int_{-\infty}^{t} h_{B\,im}(t,s)\, h_{B\,jm}(t,s)\, ds$$

Now let W be a large and positive scalar. We will work temporarily with the following approximation to $(J_B)_{i,j}$:

$$(J_B(W))_{i,j} = \sum_{m=1}^{L} \int_0^{\tau} dt \int_{-W\tau}^{t} h_{B\,im}(t,s)\, h_{B\,jm}(t,s)\, ds$$

The exponential stability of h ensures that $(J_B(W))_{i,j} \to (J_B)_{i,j}$ as $W \to \infty$.

Now

$$(J_B(W))_{ij} = \sum_{m=1}^{L} \{ \int_0^{\tau} dt \int_0^{t} h_{B\,im}(t,s)\, h_{B\,jm}(t,s) ds + \int_0^{t} dt \int_{-\tau}^{0} h_{B\,im}(t,s)\, h_{B\,jm}(t,s) ds$$
$$+ \ldots + \int_0^{t} dt \int_{-W\tau}^{-(W-1)\tau} h_{B\,im}(t,s)\, h_{B\,jm}(t,s) ds$$

By the continuity properties of $h_\alpha(t,s)$

$$(J_B(W))_{ij} = \sum_{m=1}^{L} \lim_{N\to\infty} \{ \tau^2/N^2 \sum_{u=0}^{N-1} \sum_{v=0}^{u-1} h_{B\,im}(u\tau/N, v\tau/N)\, h_{B\,jm}(u\tau/N, v\tau/N)$$
$$+ \tau^2/N^2 \sum_{u=0}^{N-1} \sum_{v=-N}^{-1} h_{B\,im}(u\tau/N, v\tau/N)\, h_{B\,jm}(u\tau/N, v\tau/N)$$
$$+ \ldots + \tau^2/N^2 \sum_{u=0}^{N-1} \sum_{v=-Wn}^{-W(N-1)-1} h_{B\,im}(u\tau/N, v\tau/N)\, h_{B\,jm}(u\tau/N, v\tau/N) \}$$
$$= \sum_{m=1}^{L} [\lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{v=-Wn}^{u-1} h_{d\,B\,im}(u,v)\, h_{d\,B\,jm}(u,v)] \qquad \text{by Lemma 3.2}$$
$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} \sum_{s=0}^{W} h_{d\,B\,im}(u,-Ns+t)\, h_{d\,B\,jm}(u,-Ns+t)$$
$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} \sum_{s=0}^{W} [\tilde{h}_{B\,im}(s,0)]_{ut}\, [\tilde{h}_{B\,jm}(s,0)]_{ut}$$
$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{s=0}^{W} \mathrm{tr}\, [\tilde{h}_{B\,im}(s)\, \tilde{h}_{B\,jm}^{T}(s)]$$

The lemma now follows immediately.

Appendix H

Proof of Theorem 5.1

With the partitions of {A,B,C} and Σ as in (5.2.6) and (5.2.9), introduce the following notation:

$$\phi_r(s)=(sI-A_r)^{-1},$$

$$\Delta_r(s)=sI-A_{22}-A_{21}\phi_r(s)A_{12},$$

$$\Xi_r(s)=A_{21}\phi_r(s)B_r+B_2,$$

$$\Gamma_r(s)=C_r\phi_r(s)A_{12}+C_2.$$

As it has been shown in Enns (1984a),

$$K - K_r = \Gamma_r \Delta_r^{-1} \Xi_r.$$

Thus,

$$\|W(K - K_r)V\|_\infty{}^2 = \max_\omega \lambda_{max} [W \Gamma_r \Delta_r^{-1} \Xi_r V V^H \Xi_r^H \Delta_r^{-H} \Gamma_r^H W^H]$$

$$= \max_{\omega} \lambda_{max} [\Delta_r^{-1} \Pi_r \Delta_r^{-H} \Theta_r],$$

where

$$\Pi_r(s) = \Xi_r(s)V(s)V^H(s)\Xi_r^H(s)$$

and

$$\Theta_r(s) = \Gamma_r^H(s)W^H(s)W(s)\Gamma_r(s).$$

Let us make the further definition:

$$\Phi(s) = (sI - A)^{-1}.$$

Then, as is shown below, the following formulae are true:

$$\Delta_r^{-1} \Pi_r$$
$$=\Sigma_2 + [0 \quad I] P_{23} \Phi_V^H C_V^T \Xi_r^H + \Delta_r^{-1} (\Sigma_2 + \Xi_r C_V \Phi_V P_{23}^T [0 \quad I]^T) \Delta_r^H,$$

$$\Delta_r^{-H} \Theta_r$$
$$=\Sigma_2 + [0 \quad I] Q_{12}^T \Phi_W B_W \Gamma_r + \Delta_r^{-H} (\Sigma_2 + \Gamma_r B_W^T \Phi_W^H Q_{12} [0 \quad I]^T) \Delta_r.$$

Indeed, from the 3–3, 2–3 and 2–2 blocks of the Lyapunov equation (5.2.8a), one can obtain:

$$P_V A_V^T + A_V P_V + B_V B_V^T = 0,$$

$$P_{23} A_V^T + A P_{23} + B C_V P_V + B D_V B_V^T = 0,$$

$$\Sigma A^T + A\Sigma + P_{23} C_V^T B^T + B C_V P_{23}^T + B D_V D_V^T B^T = 0.$$

Then, using the equations above, one can obtain:

$$\Pi_r = \Xi_r V V^H \Xi_r^H = [A_{21} \phi_r \quad I] B V V^H B^T [A_{21} \phi_r \quad I]^H$$

$$= [A_{21}\phi_r \quad I] B \{C_V\Phi_V B_V B_V^T \Phi_V^H C_V^T + D_V B_V^T \Phi_V^H C_V^T + C_V \Phi_V B_V D_V^T + D_V D_V^T\} B^T [A_{21}\phi_r \quad I]^H$$

$$= [A_{21}\phi_r \quad I] \{-[P_{23}A_V^T + AP_{23} + BC_V P_V + BC_V\Phi_V(P_V A_V^T + A_V P_V)]\Phi_V^H C_V^T B^T - BC_V\Phi_V[A_V P_{23}^T + P_{23}^T A^T + P_V C_V^T B^T] - \Sigma A^T - A\Sigma - P_{23}C_V^T B^T - BC_V P_{23}^T\} [A_{21}\phi_r \quad I]^H.$$

Then, Π_r can be rewritten as

$$\Pi_r = [A_{21}\phi_r \quad I] \{-[P_{23}A_V^T + AP_{23}]\Phi_V^H C_V^T B^T - BC_V\Phi_V[A_V P_{23}^T + P_{23}^T A^T]$$

$- \Sigma A^T - A\Sigma - P_{23}C_V{}^T B^T - BC_V P_{23}{}^T\} [A_{21}\phi_r \quad I]^H$

$= [A_{21}\phi_r \quad I] \{-[P_{23}A_V{}^T + AP_{23}]\Phi_V{}^H C_V{}^T B^T + BC_V\Phi_V P_V C_V{}^T B^T$
$- BC_V\Phi_V[A_V P_{23}{}^T + P_{23}{}^T A^T] - BC_V\Phi_V P_V C_V{}^T B^T \;\; - \Sigma A^T - A\Sigma - P_{23}C_V{}^T B^T$
$- BC_V P_{23}{}^T\} [A_{21}\phi_r \quad I]^H$,

and using the same equations above,

$\Pi_r = [A_{21}\phi_r \quad I] \{\Phi^{-1}P_{23}\Phi_V{}^H C_V{}^T B^T + BC_V\Phi_V P_{23}{}^T\Phi^{-H} - \Sigma A^T - A\Sigma\}$
$\times \; [A_{21}\phi_r \;\; I]^H$

$= \Delta_r \; \Sigma_2 + \Sigma_2 \; \Delta_r{}^H + \Xi_r \; C_V \; \Phi_V \; P_{23}{}^T \; [0 \quad \Delta_r]^H$
$+ [0 \quad \Delta_r] \; P_{23} \; \Phi_V{}^H \; C_V{}^T \; \Xi_r{}^H.$

Thus,

$\Delta_r{}^{-1} \; \Pi_r$
$=\Sigma_2 + [0 \quad I] \; P_{23} \; \Phi_V{}^H \; C_V{}^T \; \Xi_r{}^H + \Delta_r{}^{-1} \; (\Sigma_2 + \Xi_r \; C_V \; \Phi_V \; P_{23}{}^T \; [0 \quad I]^T) \; \Delta_r{}^H.$

Similarly, using the Lyapunov equation (5.2.8b), one can obtain

$$\Delta_r{}^{-H} \; \Theta_r = \Sigma_2 + [0 \quad I] \; Q_{12}{}^T \; \Phi_W \; B_W \; \Gamma_r$$
$$+ \; \Delta_r{}^{-H} \; (\Sigma_2 + \Gamma_r \; B_W{}^T \; \Phi_W{}^H \; Q_{12} \; [0 \quad I]^T) \; \Delta_r.$$

Now consider the case when the state dimension is reduced by one, i.e. $r = n-1$.

Then

$$\Delta_{n-1}{}^{-1} \; \Pi_{n-1} = (\sigma_n + \Xi_{n-1} \; C_V \; \Phi_V \; P_{23}{}^T \; [0 \; . \; . \; . \; 0 \;\; 1]^T) \; (1 + \Delta_{n-1}{}^{-1} \; \Delta_{n-1}{}^H),$$

$$\Delta_{n-1}{}^{-H} \; \Theta_{n-1} = (\sigma_n + [0 \; . \; . \; . \; 0 \;\; 1] \; Q_{12}{}^T \; \Phi_W \; B_W \; \Gamma_{n-1}) \; (1 + \Delta_{n-1}{}^{-H} \; \Delta_{n-1}),$$

and

$$\|W(K - K_{n-1})V\|_\infty{}^2$$

$$= \|(\sigma_n + \Xi_{n-1} \; C_V \; \Phi_V \; P_{23}{}^T \; [0 \; . \; . \; . \; 0 \;\; 1]^T)$$
$$\times (\sigma_n + [0 \; . \; . \; . \; 0 \;\; 1] \; Q_{12}{}^T \; \Phi_W \; B_W \; \Gamma_{n-1})$$
$$\times (1 + \Delta_{n-1}{}^{-1} \; \Delta_{n-1}{}^H) \; (1 + \Delta_{n-1}{}^{-H} \; \Delta_{n-1})\|_\infty$$

$$\leq \ 4\ (\sigma_n + \|\Xi_{n-1}\|_\infty \ \|C_V\ \Phi_V\ P_{23}{}^T\ [0 \ldots 0 \ \ 1]^T\|_\infty)$$
$$\times(\sigma_n + \|[0 \ldots 0 \ \ 1]\ Q_{12}{}^T\ \Phi_W\ B_W\|_\infty \ \|\Gamma_{n-1}\|_\infty).$$

At this point we need the following inequalities:

$$\|C_V\ \Phi_V\ P_{23}{}^T\ [0 \ldots 0 \ \ 1]^T\|_\infty \leq \sqrt{\sigma_n}\ \|C_V\ \Phi_V\ P_V{}^{1/2}\|_\infty \tag{H1}$$

$$\|[0 \ldots 0 \ \ 1]\ Q_{12}{}^T\ \Phi_W\ B_W\|_\infty \ \leq \sqrt{\sigma_n}\ \|Q_W{}^{1/2}\ \Phi_W\ B_W\|_\infty. \tag{H2}$$

To show that they are true, let us first consider the following positive definite submatrix of the gramian $\overline{Q}$:

$$\left(\begin{array}{c|ccc} Q_W & q_{12,1} & \cdots & q_{12,n} \\ \hline q_{12,1}{}^T & \sigma_1 & & \\ \vdots & & \ddots & 0 \\ \vdots & 0 & & \ddots \\ q_{12,n}{}^T & & & \sigma_n \end{array}\right) > 0$$

Then, it follows that

$$\sigma_n\ Q_W - q_{12,n}\ q_{12,n}{}^T > 0.$$

Hence, we obtain

$$\|[0 \ldots 0 \ \ 1]\ Q_{12}{}^T\ \Phi_W\ B_W\|_\infty \ = \ \|q_{12,n}{}^T\ \Phi_W\ B_W\|_\infty$$
$$\leq \ \sqrt{\sigma_n}\ \|Q_W{}^{1/2}\ \Phi_W\ B_W\|_\infty.$$

Similar derivations lead to the first inequality (H1).

According to these inequalities,

$$\|W\ (K - K_{n-1})\ V\|_\infty{}^2 \ \leq \ 4\ (\sigma_n + \alpha_n\sqrt{\sigma_n}\ )\ (\sigma_n + \beta_n\sqrt{\sigma_n}\ ),$$

where

$$\alpha_n = \|\Xi_{n-1}\|_\infty\ \|C_V\ \Phi_V\ P_V{}^{1/2}\|_\infty$$

and

$$\beta_n = \|Q_W{}^{1/2}\ \Phi_W\ B_W\|_\infty\ \|\Gamma_{n-1}\|_\infty.$$

Thus, the reduction error can be bounded as

$$\|W\,(K - K_{n-1})\,V\|_\infty \leq 2\,\sqrt{\sigma_n^2 + (\alpha_n + \beta_n)\sigma_n^{3/2} + \alpha_n\,\beta_n\,\sigma_n}\,.$$

The result can be extended to the general case of $r \leq n-1$.
In fact,

$$\|W\,(K - K_r)\,V\|_\infty = \|W\,(K - K_{n-1} + K_{n-1} - K_{n-2} + \ldots + K_{r+1} - K_r)\,V\|_\infty$$

$$= \|\sum_{k=r+1}^{n} [W\,(K_k - K_{k-1})\,V]\|_\infty \leq \sum_{k=r+1}^{n} \|W\,(K_k - K_{k-1})\,V\|_\infty.$$

As we have shown,

$$\|W\,(K_k - K_{k-1})\,V\|_\infty \leq 2\,\sqrt{\sigma_k^2 + (\alpha_k + \beta_k)\sigma_k^{3/2} + \alpha_k\,\beta_k\,\sigma_k}\,.$$

Then the main result of the theorem follows immediately. ❑

Appendix I

Proof of Theorem 6.2

As shown in Enns (1984a), consider

$K-K_{n-1} = \Gamma_{n-1}(sI-a_{nn}-a_{21}{}^{n}\ \phi_{n-1}\ a_{12}{}^{n})^{-1}\ \Xi_{n-1} = \Gamma_{n-1}\ \Delta_{n-1}{}^{-1}\ \Xi_{n-1}.$

Thus,

$$\|W(K-K_{n-1})V\|_{\infty}{}^{2}$$

$$= \max_{\omega}\ \lambda_{max}[W\Gamma_{n-1}\Delta_{n-1}{}^{-1}\Xi_{n-1}VV^{H}\Xi_{n-1}{}^{H}\Delta_{n-1}{}^{-H}\Gamma_{n-1}{}^{H}W^{H}]$$

$$= \max_{\omega}\ \lambda_{max}[\Delta_{n-1}{}^{-1}\ \Pi_{n-1}\ \Delta_{n-1}{}^{-H}\ \Theta_{n-1}]$$

where

$$\Pi_{n-1}(s) = \Xi_{n-1}(s)\ V(s)\ V^{H}(s)\ \Xi_{n-1}{}^{H}(s)$$

$$\Theta_{n-1}(s) = \Gamma_{n-1}{}^{H}(s)\ W^{H}(s)\ W(s)\ \Gamma_{n-1}(s)$$

Expanding the (1,1), (1,2) and (2,2) blocks of the Lyapunov equation (6.3.2) we get

$$AP+PA^T+BC_VP_{12}^T+P_{12}C_V^TB^T+BD_VD_V^TB^T=0$$

$$AP_{12}+P_{12}A_V^T+BC_VP_V+BD_VB_V^T=0$$

$$A_VP_V+P_VA_V^T+B_VB_V^T=0$$

Using the above equations, we obtain:

$$\Pi_{n-1} = \Xi_{n-1}VV^H\Xi_{n-1}^H = [a_{21}^n \ \phi_{n-1} \quad I]BVV^HB^T[a_{21}^n \ \phi_{n-1} \quad I]^H$$

$$= [a_{21}^n\phi_{n-1} \quad I]$$
$$\times B\{C_V\Phi_VB_VB_V^T\Phi_V^HC_V^T+D_VB_V^T\Phi_V^HC_V^T+C_V\Phi_VB_VD_V^T+D_VD_V^T\}B^T$$
$$\times [a_{21}^n\phi_{n-1} \quad I]^H$$

$$= [a_{21}^n\phi_{n-1} \quad I]$$
$$\times\{-[AP_{12}+P_{12}A_V^T+BC_VP_V+BC_V\Phi_V(A_VP_V+P_VA_V^T)]\Phi_V^HC_V^TB^T$$
$$-BC_V\Phi_V[A_VP_{12}^T+P_{12}^TA^T+P_VC_V^TB^T]-AP-PA^T-BC_VP_{12}^T-P_{12}C_V^TB^T\}$$
$$\times [a_{21}^n\phi_{n-1} \quad I]^H$$

$$= [a_{21}^n\phi_{n-1} \quad I]\ \{-[AP_{12}+P_{12}A_V^T]\Phi_V^HC_V^TB^T-BC_V\Phi_V[A_VP_{12}^T+P_{12}^TA^T]$$
$$-AP-PA^T-BC_VP_{12}^T-P_{12}C_V^TB^T\}[a_{21}^n\phi_{n-1} \quad I]^H$$

$$= [a_{21}^n\phi_{n-1} \quad I]\ \{BC_V\Phi_VP_{12}^T\Phi^{-H}+\Phi^{-1}P_{12}\Phi_V^HC_V^TB^T-A\Sigma-\Sigma A^T$$
$$+\Phi^{-1}P_{12}P_V^{-1}P_{12}^T+P_{12}P_V^{-1}P_{12}^T\Phi^{-H}\}[a_{21}^n\phi_{n-1} \quad I]^H$$

$$= \Delta_{n-1}\ \sigma_n+[0 \quad \Delta_{n-1}][P_{12}\Phi_V^HC_V^TB^T+P_{12}P_V^{-1}P_{12}^T][a_{21}^n\phi_{n-1} \quad I]^H$$
$$+\sigma_n\Delta_{n-1}^H+[a_{21}^n\phi_{n-1} \quad I][BC_V\Phi_VP_{12}^T+P_{12}P_V^{-1}P_{12}^T][0 \quad \Delta_{n-1}]^H.$$

Thus,

$$\Delta_{n-1}^{-1}\Pi_{n-1}$$

$$=\sigma_n+e_n^TP_{12}\Phi_V^HC_V^T\Xi_{n-1}^H+\Lambda_{n-1}^H+\Delta_{n-1}^{-1}(\sigma_n+\Xi_{n-1}C_V\Phi_VP_{12}^Te_n+\Lambda_{n-1})\Delta_{n-1}^H.$$

Similarly using the other Lyapunov equation (6.3.3) we can obtain

$$\Delta_{n-1}^{-1}\ \Theta_{n-1}$$
$$=$$
$$\sigma_n+e_n^TQ_{12}^T\Phi_WB_W\Gamma_{n-1}+\Omega_{n-1}+\Delta_{n-1}^{-H}(\sigma_n+\Gamma_{n-1}^HB_W^T\Phi_W^HQ_{12}e_n+\Omega_{n-1}^H)\Delta_{n-1}.$$

$$\Delta_{n-1}^{-1}\Pi_{n-1} = (\sigma_n+\Xi_{n-1}C_V\Phi_V P_{12}^T e_n^T+\Lambda_{n-1})(1+\Delta_{n-1}^{-1}\Delta_{n-1}^H)$$

$$\Delta_{n-1}^{-1}\ \Theta_{n-1} = (\sigma_n+e_n^T Q_{12}^T\Phi_W B_W\Gamma_{n-1}+\Omega_{n-1})(1+\Delta_{n-1}^{-H}\Delta_{n-1})$$

and

$$\|W(K-K_{n-1})V\|_\infty^2$$

$$= \|(\sigma_n+\Xi_{n-1}C_V\Phi_V P_{12}^T e_n^T+\Lambda_{n-1})(\sigma_n+e_n^T Q_{12}^T\Phi_W B_W\Gamma_{n-1}+\Omega_{n-1}) \times(1+\Delta_{n-1}^{-1}\Delta_{n-1}^H)(1+\Delta_{n-1}^{-H}\Delta_{n-1})\|_\infty$$

$$\leq 4(\sigma_n+\|\Xi_{n-1}\|_\infty\|C_V\Phi_V P_{12}^T e_n^T\|_\infty+\|\Lambda_{n-1}\|_\infty) \times(\sigma_n+\|e_n^T Q_{12}^T\Phi_W B_W\|_\infty\|\Gamma_{n-1}\|_\infty+\|\Omega_{n-1}\|_\infty)$$

$$\|W(s)[K(s)-K_{n-1}(s)]V(s)\|_\infty \leq 2\ \{(\sigma_n+\alpha_n+\lambda_n)(\sigma_n+\beta_n+\omega_n)\}^{1/2}$$

The result can be extended to the general case of $r \leq n-1$.

$$E_a = \|W(K-K_r)V\|_\infty = \|W(K-K_{n-1}+K_{n-1}-K_{n-2}+\ldots+K_{r+1}-K_r)V\|_\infty$$

$$= \|\sum_{k=r+1}^{n}[W(K_k-K_{k-1})V]\|_\infty \leq \sum_{k=r+1}^{n}\|W(K_k-K_{k-1})V\|_\infty$$

$$= 2\sum_{k=r+1}^{n}\{(\sigma_k+\alpha_k+\lambda_k)(\sigma_k+\beta_k+\omega_k)\}^{1/2} = E_b$$

□